2018

Computational Investigations of the Structure and Spectroscopy of Small Biomolecules in the Gas Phase

Zachary Michael Smith
College of William and Mary - Arts & Sciences, zmsacc@gmail.com

Computational Investigations of the Structure and Spectroscopy of Small Biomolecules in the Gas Phase

Zachary Michael Smith

Newport News, Virginia

Bachelor of Science, College of William & Mary, 2017

A Thesis presented to the Graduate Faculty of The College of William & Mary
in Candidacy for the Degree of
Master of Science

Department of Chemistry

College of William & Mary
August, 2018

APPROVAL PAGE

This Thesis is submitted in partial fulfillment of
the requirements for the degree of

Master of Science

Zachary Michael Smith

Approved by the Committee July 2018

Committee Chair
Garrett-Robb-Guy Professor John Poutsma, Chemistry
College of William & Mary

Chancellor Professor Christopher Abelt, Chemistry
College of William & Mary

Assistant Professor Rachel O'Brien, Chemistry
College of William & Mary

ABSTRACT

Small biomolecule systems were interrogated using infrared multiple photon dissociation (IRMPD) action spectroscopy and corresponding quantum chemical calculations with a particular focus on peptide fragmentation and protonation site preference. b_2^+ and b_3^+ fragment ions with a terminal lysine homolog residue were investigated using IRMPD in the fingerprint region (1000 cm^{-1} – 2000 cm^{-1}) and a variety of computational methods. We present the first spectroscopic confirmation of b-ion formation with a lactam structure. Infrared spectra for b_2^+ fragment system indicate the presence of a mixture of structures, though final determination will require further investigation. The b_3^+ fragment ion infrared spectra show strong support for the presence of a predominantly lactam structure. Extensive computational research in this system suggests the B3LYP method to be the most computationally efficient density functional theory method for spectral predictions. However, the inclusion of p-polarization into the basis set yielded inconclusive results and should be investigated further.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to express my appreciation to Professor Poutsma—under whose supervision this investigation was conducted—for his patience, guidance, and instructive criticism throughout the investigation. I am also indebted to Professors Abelt and O'brien for their careful reading and criticism of the manuscript. Doctors Wysocki and Martens provided instrumental insight into the nature of captured spectra, and I am very thankful for their help. I would also like to thank Professor Scheerer and the members of his group for the synthesis of novel fragment structures. I also wish to extend a warm thank you to Anwar Radwan, Melanie Berger, and Zach Hasan who synthesized the peptides used in this work. I also extend thanks to the entire Ion Lab research group, especially Amy Schienschang, for their support of this project and help in editing this document.

Support for the completion of my degree was generously provided by the National Institutes for Health under project number 1R15GM116180-01. I am thankful for the tuition support provided by the Office of Graduate Studies at the College of William & Mary. Finally, I wish to express my gratitude to the CLIO and FELIX facilities and their associated staffs for their invaluable help in ensuring this project came to fruition.

This thesis is dedicated to my family who have helped me countless times along the way and to Emma, for always pretending to care about “that laser mass-spec thing…”

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Mass spectrometry is one of the leading analytical method for proteins, peptides, and other small biomolecules. The advent of soft ionization methods—such as matrix-assisted laser desorption ionization (MALDI)[1–3] and electrospray ionization (ESI)[4,5]—has brought the ability to efficiently ionize intact biomolecules for secondary analysis. Researchers have coupled these soft ionization techniques with pre-ionization separation and tandem mass spectrometry to reliably interrogate analyte populations for chemical information.

Despite the ubiquity of mass spectrometry-based biomolecule analysis, insight is limited by the dimensionality of the experiment. Mass spectrometers only measure the mass-to-charge ratio, *m/z*, of its analytes. Tandem mass spectrometry addresses some issues of the single measurement by evolving the system through some dissociation event. This allows for *n* number of *m/z* measurements where *n* is limited by the MS^n capabilities of the mass spectrometer. However, these subsequent measurement still only measures the *m/z*. Subtle structural information, such as charge location, secondary structure, protonation or metalation site, and fragmentation pathway preference, remains beyond the scope of this measurement and requires the addition of some other dimension of analysis.

Several robust methods have been developed to characterize the structural information of gas phase ions in the mass spectrometer. Electron diffraction has been used to coax out structural information such as dihedral angles from small molecules but demonstrates a weakness in characterizing metal containing systems.[6–9] UV photodissociation has been used to characterize the structure of large proteins and protein-complexes in a top-down fashion.[10,11] Ion mobility coupled with mass spectrometry has been shown to be robust enough to characterize large protein-complex and specific

enough to differentiate between enantiomers and various electronic configuration states.[12–15] Structural characterization of small biomolecules, however, can take advantage of functional groups present in the analytes.

Vibrational excitation through infrared spectroscopy offers a more direct route to sampling the chemical and structural information of a system through coherent light absorption. Intense infrared absorbers, such as CO, NH, and OH modes, are also excellent hydrogen bond acceptors and donors. Information on these modes detail not only the primary structure, but proton locations and secondary structure as well. When infrared spectroscopy is combined with more advanced computational methods, theoretical spectra can guide the experimenters to more specific information, such as hydrogen bonding motifs and adduct moieties. Infrared spectroscopy provides the missing link in the mass spectrometric analytical toolkit for small biomolecule systems. Unfortunately, traditional infrared absorption experiments are difficult to conduct on gas phase ions.

Ions of the same charge repel each other and therefore do not generally exist in dense enough concentrations to take direct absorption measurements. The first infrared experiments on ions interrogated charged water clusters via traditional absorbance measurement, overcoming the concentration limitation through relatively large infrared cross-sections and signal averaging over long periods of time.[16] While some groups were successful in this method of ion spectroscopy, the advent of infrared multiple photon dissociation (IRMPD) spectroscopy in the latter half of the 1990s increased the ease and versatility of infrared spectroscopy on ions.

IRMPD spectroscopy avoids this dilute concentration problem by taking an incidental *m/z* measurement in place of traditional absorption. A mass-selected molecule is irradiated by a tunable, coherent infrared light source and absorbs multiple photons that

increase its energy beyond a dissociation barrier. The unimolecular dissociation of the mass-selected molecule results in a change in the m/z value of the desired molecule, introducing a new fragment m/z value detected in the mass spectrometer. The theoretical basis of IRMPD spectroscopy revolves around monitoring the unimolecular dissociation of a desired molecule in proportion to the IRMPD fragments generated from the dissociation. In effect, an indirect absorption measurement is taken wherein frequencies in which the molecule absorbs more will generate a greater proportion of fragment ions. This type of incident spectroscopy, where light causes a change in the population that is monitored in place of the change in light itself, is called action spectroscopy. These action spectroscopy measurements are much more sensitive than direct absorption measurements as a mass spectrometer can theoretically distinguish as few as a change in a single ion's *m/z* value.

The availability of powerful tunable infrared light sources has paved the way to meaningful IRMPD experiments. Powerful optical parametric oscillator/amplifier (OPO/A) benchtop laser sources became commercially viable for homemade instrumentation and free electron laser (FEL) facilities were built to provide high powered, high frequency range support for IRMPD experiments.

IRMPD is a radiative dissociation as a result of the absorption of numerous photons of discreet energy. IRMPD is dependent on the absorption of multiple photons and cannot therefore be thought of as a direct function of absorption cross-section. Instead, IRMPD should be thought of as a combination of individual absorption events that raise the internal energy of a molecule beyond a vibrational dissociation barrier (see **Figure 1**). The increase in the internal energy results in fragmentation along the weakest bond in the molecule as opposed to the vibrationally-excited mode.

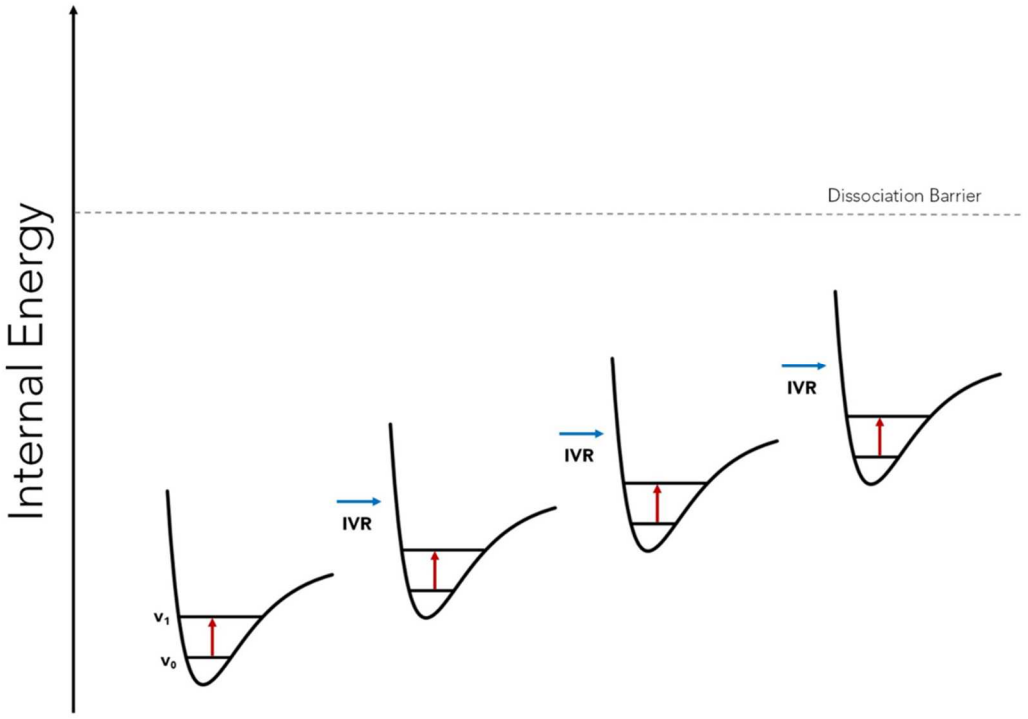


Figure 1 This graphic diagrams the excitation of a molecule from v_0 to v_1 as the result of absorption of photon. Immediately following the absorption of the photon, the molecule undergoes intramolecular vibrational relaxation (IVR) which redistributes the vibrational energy amongst other vibrational modes equally, increasing the internal energy of the molecule. IVR also reopens the initial excited mode for reabsorption. This process continues until the dissociation threshold is surpassed.

As the radiative event is not the direct cause of cleavage, the multiple-photon dissociation pathway is understood to proceed through a non-coherent mechanism.[17] After the molecule is excited along a specific vibrational mode, the energy quickly dissipates among the various vibrational degrees of freedom in a process called intramolecular vibrational relaxation (IVR).[18] IVR reopens the incident vibrational excitation mode, $v_0 \rightarrow v_1$, to reabsorb and continue the multiple photon absorption process. IVR increases the overall internal energy of the molecule by redistributing the excitation energy and results in a unimolecular dissociation once the internal energy of the molecule surpasses the dissociation threshold. IRMPD is therefore not a state-specific dissociation pathway—

rather than dissociating along the vibrationally-excited mode, the molecule dissociates along the lowest energy bond.

The IRMPD process is similar to that of CID in an ion trap mass spectrometer. In an ion trap, ions are submerged in a bath of helium at constant pressure. CID activation uses radio frequency (RF) excitation to increase the oscillation amplitude of selected masses. The RF excitation reduces the mean free path distance of the mass-selected ion, thereby increasing the rate and energy at which the ion collides with a helium molecule. The higher rate of collision heats the mass-selected ion, ultimately increasing the ion's internal energy beyond a dissociation threshold. IRMPD, however, can discriminate further beyond mass-selection through the use of a tunable light source. For successful IRMPD action spectroscopy experiments the focus is not on the type of fragment produced—IRMPD will result in similar fragments across the entirety of the measured spectrum—but instead on the intensity of the fragment signal that is generated at a specific frequency.

The weaknesses of IRMPD spectroscopy come from the loss of orthogonality at higher internal energy states, non-dissociative losses in energy, and collisional cooling. The IRMPD process avoids the anharmonicity bottleneck of repeated stepwise excitation along the same vibrational mode through IVR.[19] However, at higher internal energies, the normal excitation modes lose their orthogonality. This results in slight red-shifting of bands associated with higher dissociation barriers and loss of laser frequency-absorption overlap. This effect is mitigated in trapped mass spectrometers by the increased absorption time window and the quasi-continuum effect wherein ions at higher internal energy present a broadening of their absorption cross-section.[19] While absorption cross-section widening tempers the red-shifted absorption window, it also contributes to spectral broadening during the experiment.

Non-dissociative energy losses present themselves as spontaneous and stimulated emissions. Spontaneous emission is the radiative decay of the excited electron without interference from an external source. As this process is spontaneous, this will factor into IRMPD efficiency. Stimulated emissions, the decay of an excited electron due to further radiation, is not a serious detractor to IRMPD efficiency. The increased density of states at higher internal energy levels leads to degeneracy at the excited state, encouraging further absorption over stimulated emission.

Trapping mass spectrometers use collisional cooling to reduce the momentum of incoming ions. Collisional cooling reduces momentum by transferring internal energy from the trapped ions to an inert cooling gas, like helium. The energy loss associated with collisional cooling can compete with the completion of the IRMPD pathway. Though this competition can suppress IRMPD efficiency, the effect should present itself equally across wavelengths and is combatted by sufficiently powerful lasing sources. Competition from collisional cooling will be mitigated by a collision-less environment produced under high vacuum ($< 10^{-7}$ mbar), such as those present in an ICR.[20] These non-dissociative energy losses can affect IRMPD efficiency, effectively resulting in uniform dissociation suppression. As discussed, these effects can be mitigated and will therefore be largely negligible. However, these losses in energy should always be considered when analyzing spectra, especially those with inefficient fragmentation or low signal:noise ratios.

IRMPD has been used to characterize peptide structure,[21–23] metal-adduct moiety,[24,25] protonation site preference,[26,27] cluster formation,[28–30] and isomer differentiation.[12,31,32] IRMPD action spectroscopy has carved out its niche in the mass spectrometry community as a powerful method for structural identification, especially in small biomolecule systems. This work coupled IRMPD spectroscopy with robust computational methods to describe various biologically relevant systems.

Chapter 2: Mass Spectrometry-Based Proteomics

The biological system is dependent on cells—self-regulating, autonomous units whose complex genetic programming can be simplified to a directive to maintain and pass on information. This information is maintained in the DNA and is translated into proteins—biomolecules of varying size responsible for a large majority of biological functions—by way of transcribed RNA. The first large-scale, high-throughput attempt to synthesize this biological information in a usable manner was genomics. Genomics is the mapping and editing of the structure and function of genomes—the complete set of an organism's DNA.[33] Included in this mapping is the full set of available genes, a template for the proteins an organism can produce as a result from gene expression.

However, genes and proteins do not have a one-to-one relationship. Simply because an organism can produce a protein does not mean it will, and if it does, it certainly will not produce every protein in the same amount. Further, proteins do not stay at the same location they are made. Proteins move about and throughout the cell to perform its function. As such, proteins are susceptible to changes in their microenvironment. Physical changes—such as tight and loose conformations—can change in response to the change in the protein's microenvironment. As the protein carries out its function, chemical changes—such as binding to a receptor or other post-translational modifications—can affect subsets of the protein population. Genes, however, cannot communicate these changes.

In response to the desire to characterize biological information more precisely, the field of proteomics was born. Proteomics is the large-scale analysis of proteins for their structure and function.[34] Proteomics move beyond genomics into the differential realm, allowing for a "real time" snapshot of a cell's proteome. If genomics, a map of what a cell

could possibly do, is trigonometry, then proteomics, a picture of what the cell is presently doing, is calculus.

Mass spectrometry-based proteomics has increased the ease with which proteins can be identified and sequenced for biological research. Each amino acid, apart from leucine and isoleucine, has a unique weight. Unique, known weights in a mass spectrometer mean that sequencing can occur via some form of tandem mass spectrometry evolution, primarily collision induced dissociation. The combination of liquid chromatography separation and mass spectrometry in time allows for high-throughput analysis of a cell's proteome. There are two common categories of mass spectrometry-based proteomics experiments: top-down proteomics and bottom-up proteomics.

Top down proteomics experiments differ from bottom-up experiments in their treatment of the purified protein. Top-down methods inject an intact protein into the mass spectrometer for analysis. While methods differ in terms of pre-mass spectrometry protein treatment, there is generally some form of separation beforehand so that only one protein is being analyzed at a time.[35] Once in the mass spectrometer, the protein will be activated and fragmented. Computational software is then used to identify the fragments and assign peptide identities to them.

Top-down proteomics has two large limitations in its applicability. The first comes from the intact protein's size. Most mass spectrometers have a relatively small upper mass range. Ion traps are typically ineffective past 4000 *m/z* in high range mode and quadrupole instruments have an even lower limit of 1500 *m/z*. The second limitation comes from resolving power of the mass spectrometer. Intact proteins will ionize at a range of charge states, and the ^{13}C peaks from a molecule of significant size will broaden each charge state over a range of measured *m/z*. Unless the mass spectrometer has sufficient resolving power to differentiate these broadened peaks, direct analysis of intact proteins

will be nearly impossible. A time of flight mass analyzer, for instance, will solve both limitations but also limits the experiment to MS^2 in most cases and requires a relatively high-end instrument.

In a bottom-up proteomics approach, instead, one breaks proteins into smaller peptides before analysis. Following protein purification, the protein is digested by some lysing enzyme to destroy any secondary and tertiary structure and fractionate the protein into enzyme specific peptides. The peptide mixture is then injected into an HPLC for pre-mass spectrometer separation. This reduces the sample received at the mass spectrometer for increased signal of the individual peptides. The peptides are fragmented in the mass spectrometer, and then bioinformatics software identifies peptides from these fragments.

Bottom-up proteomics avoids both limitations seen in the top-down methods. Peptide size is dependent on the protein and lysing enzyme used but will generally range from 10 to 15 residues in length. Peptides of this length are well within the range of most mass spectrometers especially if the peptide is doubly or triply protonated. Further, with these sized peptides, nominal mass resolution is sufficient to differentiate peptide isotopomers. Bottom-up proteomics also has the benefit of increased separation from the liquid chromatography step after purification and lysis. This typically results in better signal and higher accuracy. The most notable downside from the bottom-up method is the loss of de novo sequencing capability except under rare circumstances.[36] Structural information is thrown away during the protein lysing step which also increases the difficulty of sequencing a sample with intact PTMs.

Arguably the most important component of the proteomics experiment is the bioinformatics software used for peptide analysis. Mass spectrometric data is collected for the precursor ion and its CID product ions corresponding to each proteolytic peptide and

its fragments. Typical software, such as SEQUEST, has the expected proteins for a given sample preloaded into a .fasta file. The software loads the precursor data and generates theoretical fragmentation data from all permutations of residues that could yield that precursor weight. The software then compares its expected fragmentation spectra with the observed spectrum. The peptide is assigned by the theoretical data that fits the observed best, and the assignment receives a correlation value to reflect the confidence of the assignment.

Accurate identification of proteins relies on the instrument's ability to efficiently fragment the peptides and the software's ability to correctly predict the expected fragmentation patterns. Understandably, the software needs some way to know how these peptides could fragment and in what relative intensity they will fragment. It has been shown that fragmentation mechanisms are dependent on the residues along which the peptide is fragmenting.[37] On method for detailing fragmentation pathway preference comes from working backwards from a known product structure, the fragment. Therefore, understanding the structures of the products and the fragmentation mechanisms acting to produce those products is necessary to improve mass spectrometry based-proteomics.

Gas-phase peptide fragmentation occurs along the amide backbone of the peptide (See **Figure 2**). a-, b, and c-ions are formed from dissociation with the charge remaining on the N-terminal side of the peptide while x-, y-, and z-ions are formed from dissociation with the charge remaining on the C-terminal side of the peptide. The site of fragmentation is dependent on the type of dissociation event. a- and x-ions are formed as secondary fragments from b- and y-ions or from high energy collisions in sector instruments.[38] c- and z-ions are formed via radical intermediates and therefore require specialized fragmentation conditions, such as electron capture/transfer dissociation.[39] b- and y-ion products predominate vibrational dissociation and low-energy CID methods in peptides.

Due to the high number of vibrational dissociation methods available and the ubiquity of their implementation in mass spectrometers, the analytical usefulness of b- and y-ions significantly outweighs that of the other fragment ions.

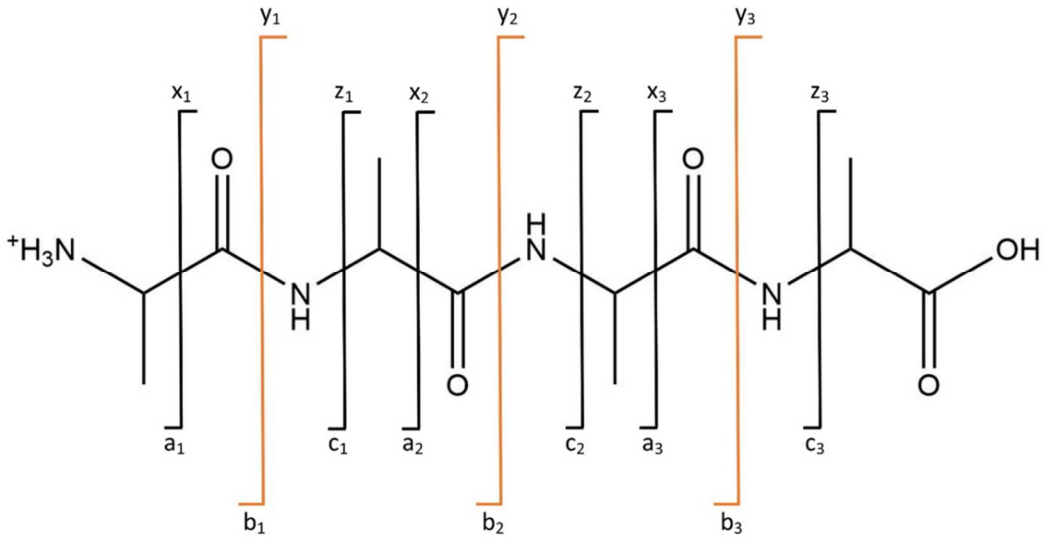


Figure 2 Peptide cleavage locations shown along the peptide backbone. Arrow direction indicates the protonated portion of the fragment. b- and y-ions highlighted as the predominate product discussed in this work.

The nomenclature developed for peptide fragmentation is f_n^+ where f is the type of fragment (a, b, c, x, y, or z) and n is the location of the fragment along the peptide in residues. Residue numbering for the fragment starts from the N-terminus for a-, b-, and c-ions while numbering for x-, y-, and z- ions starts from the C-terminus. The fragmentation sites of the peptide are governed by the mobile proton model of peptide fragmentation.[40] With few exceptions, such as those due to residue specific localization, protons are not sequestered at the most basic site of the ion and can instead populate other protonation sites upon activation to yield excellent fragmentation efficiency over a wide range of applied energies.[41] This increased fragmentation efficiency allows for cleavage at different sites in the peptide and paves the way to predictive, reproducible fragmentation patterns

for peptides. While all fragmentation pathways are not populated evenly, most are sampled with analytically relevant yields.

The smallest commonly occurring b-ion, b_2^+, is an ideal system to monitor the effects of residue location on the formation of a b-ion fragment. Previous studies have shown two principal fragmentation pathways prevail in the formation of b_2^+ ion: an oxazolone structure or a diketopiperazine structure (see **Figure 3**). The oxazolone b_2^+ is characterized by a five-membered resulting from a peptide carbonyl O's nucleophilic attack on the next carbonyl C and is resolved by proton transfer and dissociation. This pathway retains the *trans* conformation from the precursor peptide. The diketopiperazine structure is characterized by a six-membered ring formed from the N-terminus's nucleophilic addition to the second residue's carbonyl C and is resolved by proton transfer and dissociation at the second residue's carbonyl. The diketopiperazine pathway is contingent on the *trans-cis* isomerization of the first residue's amide bond.[42] Despite the diketopiperazine being more thermodynamically stable for most residues, the oxazolone form is found more commonly, likely due to the increased barrier to formation presented in a *trans-cis* isomerization or in the ring-closure step.[43]

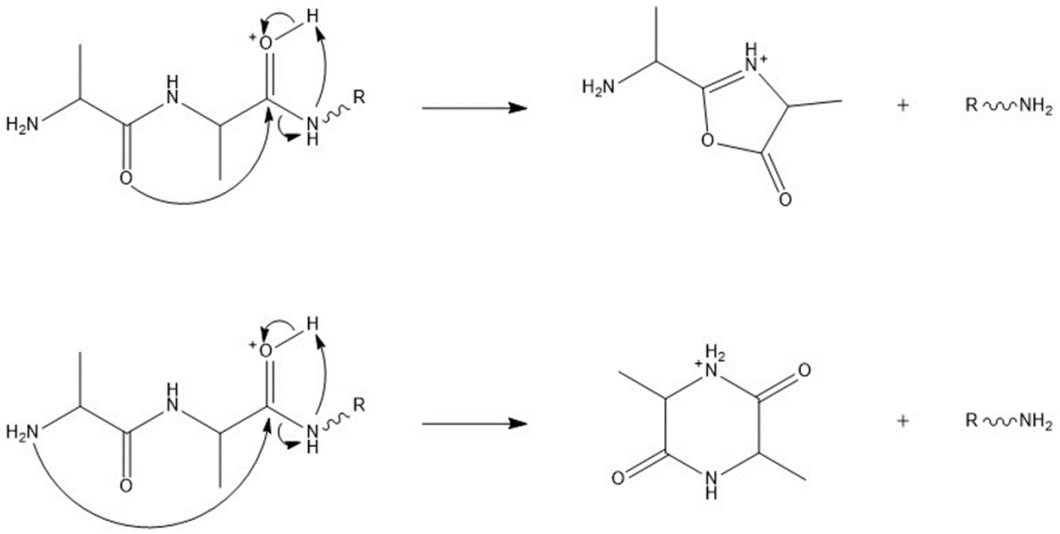


Figure 3 The fragmentation mechanisms for the oxazolone and diketopiperazine forms of the b_2^+ ion are shown. Though these mechanisms are shown in a condensed manner for ease of reading, the mechanisms are not concerted and instead proceed in a stepwise manner.

Additional fragmentation pathways have been suggested when specific residues are present. A lactam b-ion pathway has been theorized for residues with an amino side chain at the cleavage position.[44] The fragmentation pathway (See **Figure 4**) proceeds via nucleophilic attack of the amino side chain at its own carbonyl group, casting off a neutral y-fragment structure or water in the case of the amino residue positioned at the C-terminus. The lactam ring size is dependent on the length of side chain and retains the *trans* conformation of the peptide. The mechanism was theorized for protein amino acids—such as lysine, arginine, and histidine—but has been shown to function with ornithine, a non-protein amino acid homolog of lysine.

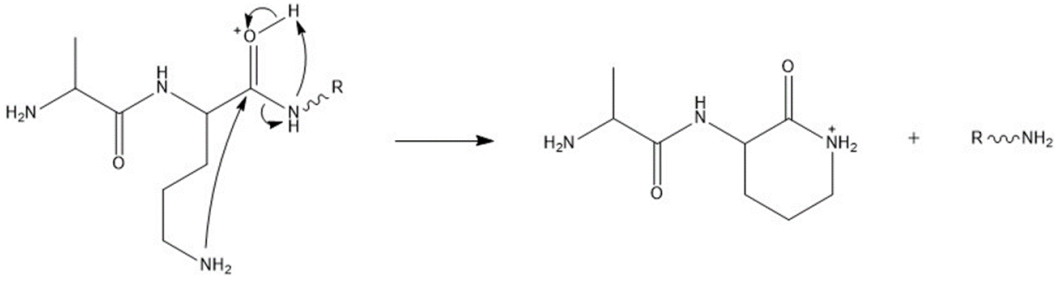


Figure 4 The fragmentation mechanisms for the lactam form of the b_2^+ ion are shown. Though this mechanism is shown in a condensed manner for ease of reading, the mechanism is not concerted and instead proceeds in a stepwise manner.

It has been shown that peptides containing an ornithine residue preferentially fragment to b-ions with ornithine in the terminal position, and that those b-ions produce fragmentation patterns consistent with that of a 6-membered lactam ring structure.[45] This preferential b-ion formation has been dubbed the “ornithine effect” and has been exploited to probe cyclic and stapled peptides in the gas phase.[46] Both solution and gas phase methods have been proposed for removal of the guanidino group of arginine to make ornithine for peptide analysis.[45,47,48] Whereas the conversion of arginine to ornithine for peptide analysis is shown to be analytically useful, confirmation of the fragmentation mechanism will develop a better understanding of the significance of the ornithine effect.

Lysine (**K**) does not show the same fragmentation preference as ornithine (**O**). **K** is an amino acid with a labile side chain consisting of a 4 methylene groups terminating in an amino group. **K** is a basic residue, previous threshold studies have placed its proton affinity (PA) at 996 kJ/mol, more basic than histidine (PA = 988 kJ/mol) but less basic than arginine (PA = 1051 kJ/mol).[49] Subsequent work has re-measured K (PA = 1004.2 kJ/mol) in comparison to **O** (PA = 1001.1) and its two shorter homologs, 2,4-diaminobutanoic acid (**B**) (PA = 975.8) and 2,3-diaminopropionoic acid (**Z**) (PA = 950.2).[50] **K**, **O**, **B**, and **Z** will be the focus of this work, and their structures are shown below (See **Figure 5**).

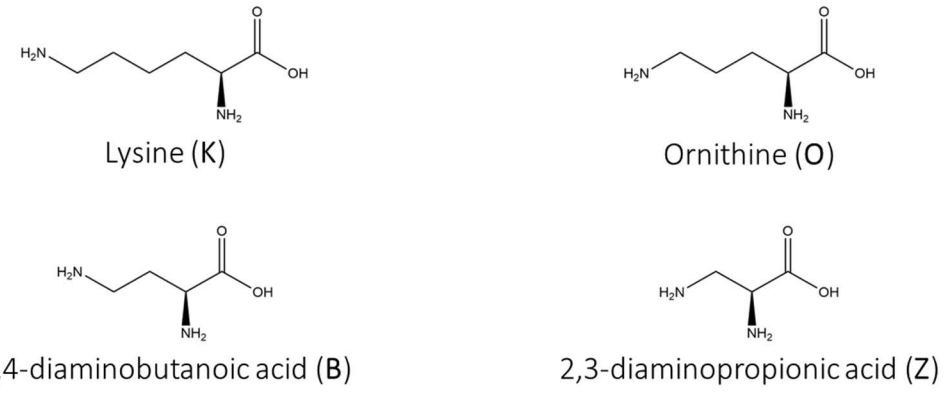


Figure 5 The structures of Lysine (**K**) and its three shorter homologs, Ornithine (**O**), 2,4-diaminobutanoic acid (**B**), and 2,3-diaminopropionoic acid (**Z**) are shown above.

The difference in PA between the three molecules is interesting. The stepwise reduction of methylene groups in the side chain implies a similarly linear reduction in proton affinity. However, **K** and **O** share similar PA (ΔPA = 3.1 kJ/mol) while **B** and **Z** present the expected reduction in PA (ΔPA = 28.4 and 54.0 kJ/mol, respectively). The small decrease in **O**'s PA is due to positive effects from entropy. Protonated **O** forms a favorable "6-membered ring" hydrogen bonding scheme between the side chain amino group and the N-terminus. **K**'s increased side chain length makes this hydrogen bonding scheme less entropically favorable (7-membered ring vs. 6-membered ring), leading to a slight increase in PA instead of the jump one might expect. Conversely, the **B** and **Z** side chains are too short to take advantage of this favorable hydrogen bonding scheme, leading to decreases in PA as a result. While **B** and **Z** do not have the biological relevance of **K** or even **O**, their structural similarity provide an excellent system to test entropy-driven change in gas-phase properties.

Entropy plays a large role in determination of gas-phase energetics. The preferential fragmentation observed from the ornithine effect suggests an entropy driven mechanism less favorable to **K**. We present a study to test the validity of that characterization. A combined computational and infrared multiple photon dissociation action spectroscopy study has been undertaken on the b_2^+ and b_3^+ systems of terminal **K**, **O**, **B**, and **Z** fragment ions. We have determined the relative stabilities and spectra of three fragment isomers to determine pathway preference and show that **O** is not the only residue capable of forming a favorable lactam structure.

Chapter 3: Basis Set and Method Comparison

Methods

Theoretical Procedures Calculations for the lactam, diketopiperazine, and oxazolone forms of AK^+, AO^+, AB^+, and AZ^+ were computed using the Gaussian 09 suite.[51] Conformer sets were generated for several protonation sites (see **Figure 6**) using the GMMX searching routine in PCModel 9[52] which varies bond lengths and angles, saving structures with energies within 40 kJ/mol of the minimum energy structure. These conformer sets were imported into Gaussian and used as starting structures for a series of molecular orbital and density functional theory calculations. Final geometries, zero-point energies, and thermal corrections to enthalpy and free energies were computed at the B3LYP/6-31+G(d)[53–55] level of theory for all species.

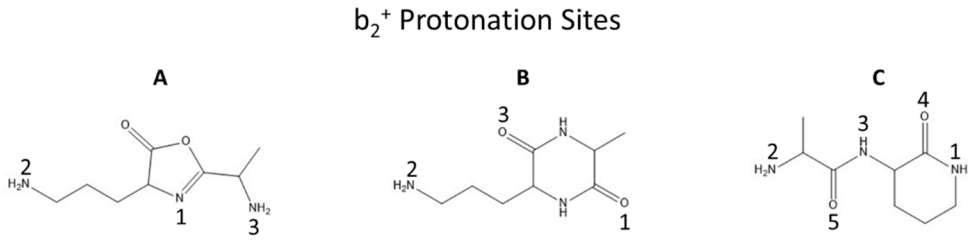


Figure 6 The b_2^+ fragment ions with ornithine in the terminal position are shown. **A** shows the oxazolone forms, **B** shows the diketopiperazine form, and **C** shows the lactam forms. The protonation sites' numbering does not imply an order of significance, rather just a naming scheme developed to differentiate conformers of the same isomer.

Naming of the conformer types follow the convention of the isomer name followed directly by a number corresponding to the protonation site (e.g. **lac4**). Diketopiperazine has been shortened to **diket**, oxazolone has been shortened to **oxa**, and lactam has been shortened to **lac**. The full word will be used when referring the isomer type while the shorthand will be used while discussing a specific conformer. The protonation site will use the naming scheme corresponding to the fragment type.

The recommended scaling factors for B3LYP frequencies are 0.98 for ZPE and 0.99 for thermal corrections to enthalpy and free energy.[56] Due to the small differences between the recommended scaling factors and the unscaled harmonic frequencies, we chose to use the unscaled harmonic frequencies to derive ZPE and thermal corrections. The reported energy stabilities were derived by extracting the total electronic energy from a ZPE computation at the B3LYP/6-311++G(d,p)[57] level of theory and converting to the 298K enthalpies and free energies using the B3LYP/6-31+G(d) computed thermal corrections. Select low-lying conformers were re-optimized at the B3LYP/6-31+G(d,p) level of theory for spectral comparison.

The B3LYP method has been shown to overestimate the effects of non-covalent interactions in computations.[58] To compare B3LYP's ability to characterize the hydrogen bonding in these systems, additional methods were investigated. The lowest lying conformer of each b_2^+ isomer was re-optimized from the PCModel starting geometry at the M06-2X/6-311++G(d,p)[59] and ωB97-XD/6-311++G(d,p)[60] levels of theory for comparison of relative stabilities and computed spectra. When geometry optimization resulted in different conformers between the three methods, additional structures were optimized to account for the missing conformer or hydrogen bonding scheme.

The NIST recommended scaling factor for the B3LYP/6-31+G(d,p) level of theory is 0.961.[61] However, both B3LYP/6-31+G(d) and B3LYP/6-31+G(d,p) spectra are unscaled for their comparison. A scaling factor of 0.955 was used for both basis sets in comparisons with the experimental data and was found by matching diagnostic peaks with in the theoretical spectra with their experimental counterparts. Neither M06-2X/6-311++G(d,p) and ωB97-XD/6-311++G(d,p) levels of theory have NIST recommended scaling factors. Therefore, scaling factors were chosen by matching diagnostic peaks to the already scaled B3LYP calculated spectra. This resulted in scaling factors of 0.924 and

0.935 for the M06-2X and ωB97-XD computed spectra, respectively. Stick spectra for all evaluated methods were broadened by a 20 cm^{-1} Gaussian function.

Results and Discussion

Basis Set Comparison The spectral differences between the B3LYP/6-31+G(d) and the B3LYP/6-31+G(d,p) levels of theory were explored. Our lab's standard computational method for determining thermochemical properties is B3LYP/6-311+G(d,p)//B3LYP/6-31+G(d). This method has shown almost quantitative agreement for thermochemical properties of amino acids and small peptides. Unfortunately, there is no recommended scaling factor for frequencies obtained using this method. In this study, we re-optimized the lowest energy structures for each isomeric cation to include the p polarization functions due to the availability of a recommended anharmonicity scaling factor for B3LYP/6-31+G(d,p). The NIST scaling factors are obtained from comparison of gas-phase vibrational spectra with calculated spectra for neutral molecules. The multiple-photon absorption process does not necessarily correspond to the single-photon Beer's law experiment. Therefore, scaling factors are decided by matching diagnostic peaks in the theoretical spectra to those present in the experimental spectra. This method of scaling factor determination often results in a scale factor that deviates from the NIST value. In order to determine if re-optimization to the larger basis set was necessary to get reliable IR spectra, a small study was conducted to determine how the addition of p-orbital polarization to the computation changes the spectra. In addition, we took this opportunity to examine other modern density functionals to determine if B3LYP is still the best choice for calculating vibration spectra.

Re-optimization of the structures at the B3LYP/6-31+G(d,p) level of theory did not result in any significant structural changes to the isomers. Importantly, each re-optimized structure maintained the starting conformer type. Some X-H and hydrogen bond lengths

shifted as the p-polarization only affects hydrogen atoms. However, those shifts were small (< 0.01 Å) and their effects should be negligible. This means that any spectral differences observed will be primarily due to how p-polarization affects the method's ability to predict spectra at this basis set.

The diagnostic carbonyl stretching mode of the oxazolone isomer of AK^+ was used for comparison between the spectra computed at the B3LYP/6-31+G(d) and B3LYP/6-31+G(d,p) levels of theory. An overlay of the two computed spectra was used to assess the need for a relative scaling factor (See **Figure 7**). Without any scaling factor, the two carbonyl bands' peaks are within 8 cm^{-1} of each other. As the spectra reduce in energy from the intense carbonyl absorbance, the spectral features present in the B3LYP/6-31+G(d,p) computed spectrum present slight red-shifting relative to its B3LYP/6-31+G(d) computed counterpart. Despite this minor offset, comparison of the other isomers was performed without any wavelength scaling.

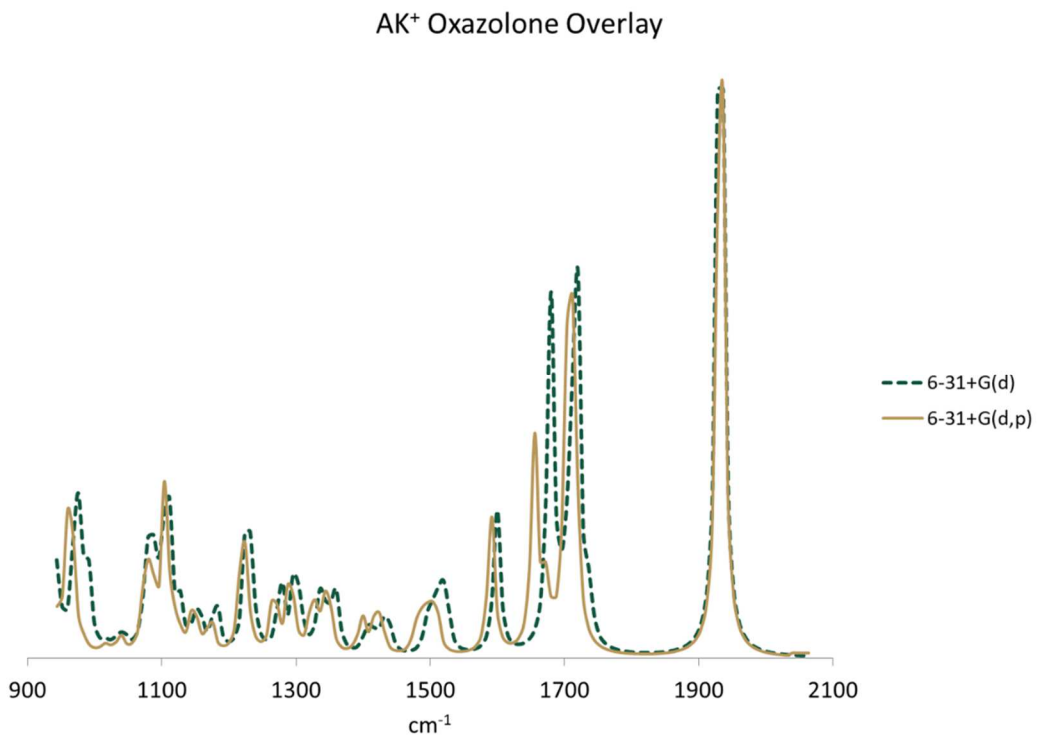


Figure 7 Spectral overlay of the oxazolone isomer of AK+ as computed at the B3LYP/6-31+G(d) (Green Dashed) and B3LYP/6-31+G(d,p) (Gold Solid) levels of theory.

Comparison of the spectra for all three AK^+ isomers reveal that most spectral features are maintained between the two calculations, though some bands have red-shifted relative to the B3LYP/6-31+G(d) computation (See **Figure 8**). The only difference immediately noticeable is the sharp, intense peak present near 1945 cm^{-1} in the B3LYP/6-31+G(d,p) optimized lactam spectrum. This peak initially looks like a diagnostic carbonyl stretching mode. However, this isomer does not have an oxygen atom in the α-position to the carbonyl to blue-shift that absorbance. Instead, this peak is the activation of the proton-bridged hydrogen bond. The lowest energy lactam isomer sequesters a hydrogen between the two carbonyls, and this absorbance is the stretching mode that oscillates between the two carbonyls. This mode is significantly blue-shifted (~300 cm^{-1}) in the B3LYP/6-31+G(d) optimized spectrum.

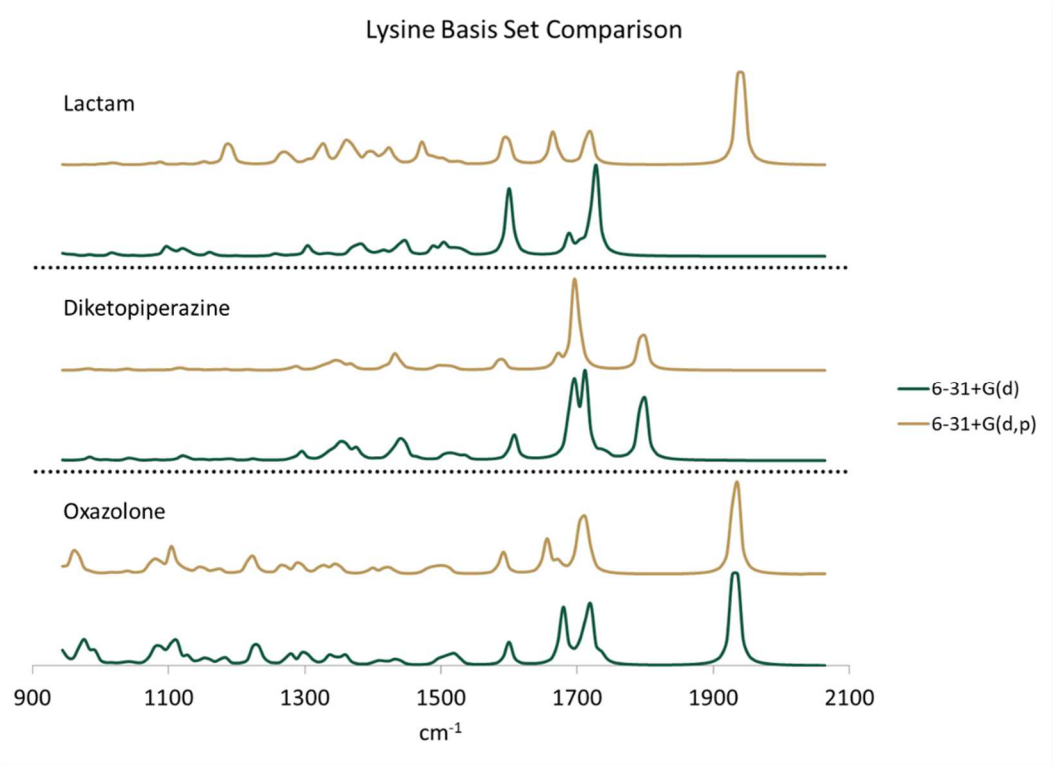


Figure 8 Full comparison of the B3LYP/6-31+G(d) (Green) and B3LYP/6-31+G(d,p) (Gold) computed spectra for the three isomers of AK^+.

As expected, the carbonyls of each of the isomers have not shifted. However, the p-polarization has affected the treatment of the N-H absorption modes. This effect is shown in the B3LYP/6-31+G(d,p) lactam spectrum where the NH_2 bend is red-shifted by ~25 cm^{-1}. This effect also presents itself to a lesser extent in the oxazolone spectrum where the amino bend is red-shifted by ~25 cm^{-1}. Despite the slight red-shifting of the NH_2 bending modes, the intensities of spectral features are largely preserved between both spectra of each isomer. Though the intensities look significantly different in the lactam spectra, the relative reduction in intensity for the B3LYP/6-31+G(d,p) spectrum results from the inclusion of the proton bridge absorbance into the normalization. Were the same peak included in the other spectrum, the two would show similar intensities.

A small shoulder present in the B3LYP/6-31+G(d,p) oxazolone spectrum due to red-shifting of the amino bending doublet near 1680 cm^{-1} is not present in the B3LYP/6-

31+G(d) doublet. In the B3LYP/6-31+G(d) spectrum, the amino bend and carbonyl stretch present as a doublet at 1696 cm^{-1} and 1728 cm^{-1} with the protonated N-H bending out of plane with respect to the carbonyl in different directions for each doublet peak. The same carbonyl stretch and amino bend present themselves as a single intense peak at 1698 cm^{-1} with a lower energy shoulder at 1680 cm^{-1} in the B3LYP/6-31+G(d,p) spectrum. The change in intensity comes from the protonated N-H bending mode staying in plane with respect to the carbonyl with the shoulder representing the small out of plane bending mode.

The spectra for AO^+ show nearly identical treatment in the predicted spectra (See **Figure 9**). Some slight red shifting of N-H modes occurs in the B3LYP/6-31+G(d,p) calculated spectra, such as the NH_2 appearing ~20 cm^{-1} lower in energy for both the lactam and diketopiperazine isomers. However, the main offender remains the sharp proton bridge absorbance near 2000 cm^{-1}. Similar to the AK^+ experimental spectra, the AO^+ experimental spectrum does not show any absorbance in that range.

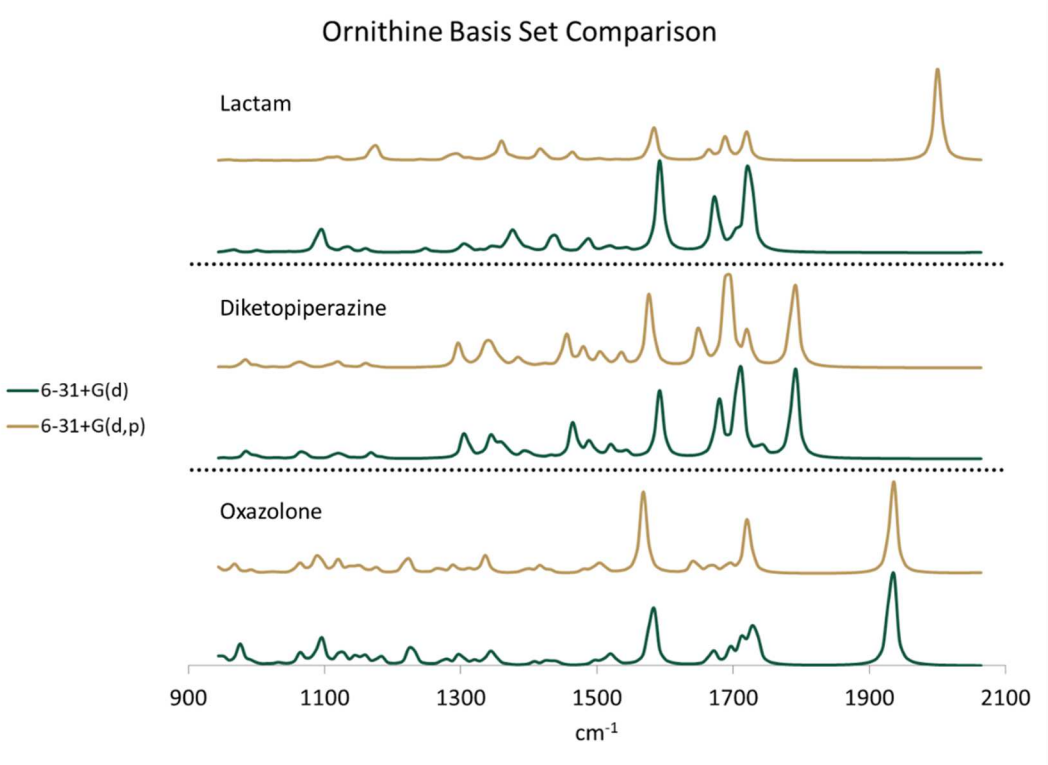


Figure 9 Full comparison of the B3LYP/6-31+G(d) (Green) and B3LYP/6-31+G(d,p) (Gold) computed spectra for the three isomers of AO^+.

AB^+ also predicts the lowest energy lactam to protonate at the alanine carbonyl, forming a proton bridge with the lactam ring carbonyl. However, calculations with both basis sets place that absorbance outside of the observed window. As such, the spectra look similar for each of the three isomers (See **Figure 10**). Some previously noted N-H absorbance mode red shifts present themselves here as well. The B3LYP/6-31+G(d,p) lactam and diketopiperazine spectra both predict NH_2 bend red-shifting of ~25 cm^{-1}. The B3LYP/6-31+G(d,p) AB^+ lactam isomer shows a distinct second peak at 1752 cm^{-1} corresponding to the ring carbonyl that is a less pronounced shoulder of the more intense alanine residue peptide carbonyl stretching mode on the B3LYP/6-31+G(d) spectrum. This is likely an artifact from the applied 20 cm^{-1} Gaussian broadening function.

The red-shifting of the NH_3 bending mode is well explained by the change in the basis set. While d-polarization affects all non-hydrogen atoms, the addition of p-

polarization only affects hydrogen atoms. The inclusion of polarizability will help to better define the hydrogen interaction, especially at the protonation site. If the polarization results in an increase in bond length, the frequency of the vibrational mode will decrease. A decrease in frequency results in a reduction in energy, explaining the red shift observed from the addition of p-polarization.

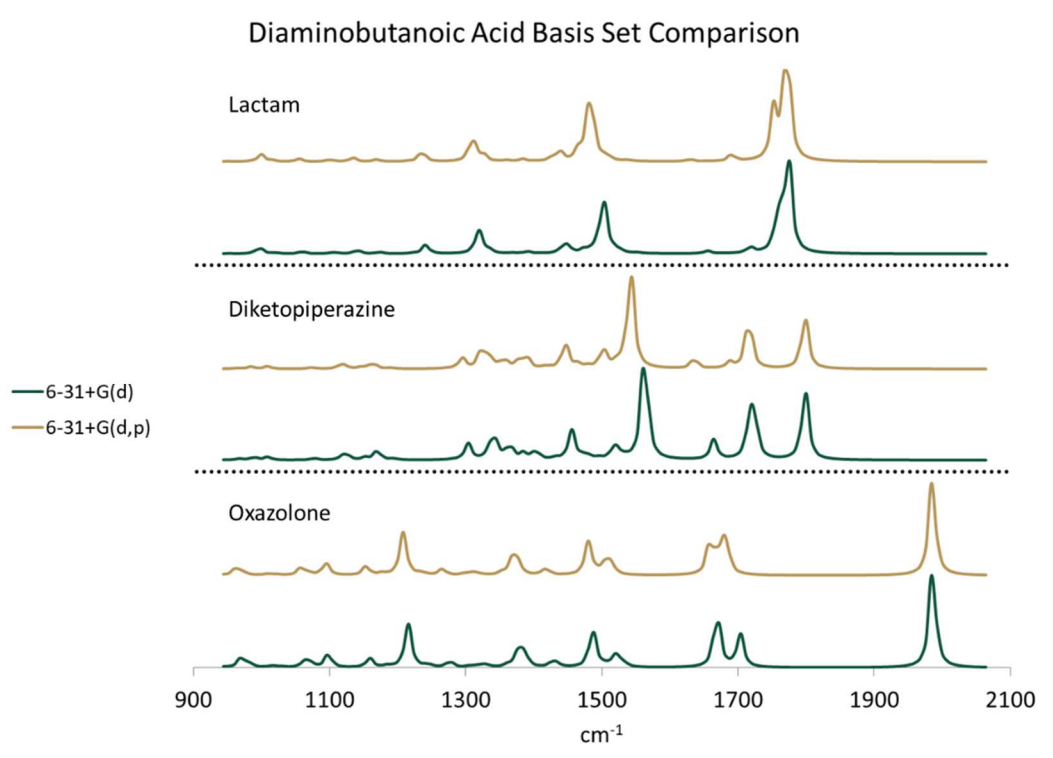


Figure 10 Full comparison of the B3LYP/6-31+G(d) (Green) and B3LYP/6-31+G(d,p) (Gold) computed spectra for the three isomers of AB^+.

While the re-optimization of these structures did not change the conformer type, re-optimization can cause a change in structure, such as the shifting of a proton from one site to another. In addition to the potential for structural changes, re-optimization changes the thermal corrections to enthalpy and free energy. These changes are not uniform across the re-optimization and instead depend on the isomer and geometry. In the case of AK^+, re-optimization raised the lowest diketopiperazine structure's free energy by 3.7kJ/mol while reducing the lactam and oxazolone forms' free energies by 3.4 and 3.8

kJ/mol, respectively. This has the net effect of reducing the diketopiperazine-lactam energy gap from 3.8 kJ/mol to -3.3 kJ/mol. This change flips the stability gap so that the lactam form is more stable than the diketopiperazine. Spectra alone cannot communicate changes in energies and stabilities. Therefore, short of re-optimization across the entire conformer set, separate basis sets should not be used for stability and spectral representations.

We will proceed with 6-31+G(d) optimized geometries and calculated spectra for the purposes of this work. The presence of intense proton bridge absorptions in the AK^+ and AO^+ spectra offset any accuracy gained from the addition of p-polarization. The severity of spectral intensity differences is overshadowed by the power correction fitting necessary to analyze the experimental data. Further, the broadening exhibited in experimental IRMPD spectra reduces the reliance on overall spectral features in favor of diagnostic bands, such as the carbonyl or amide stretching modes. The consistency of the diagnostic carbonyl peaks between the two geometry optimizations reduces the impact of the less diagnostic amino groups' red-shifting. Lastly, the NIST recommended scaling factor for the 6-31+G(d,p) basis set is only instructive as a baseline and requires refinement based on the captured experimental spectrum. This scaling refinement can be performed from the outset which eliminates the need for a recommended scaling factor thereby removing the benefit from increased calculation time.

Method Comparison The M06-2X and ωB97-XD methods were identified in previous studies for their treatment of non-covalent interactions.[58] M06-2X, like B3LYP, is a hybrid density functional which incorporates some of the Hartree-Fock energy into its calculation. The Minnesota functional doubles the contribution of Hartree-Fock exchange as compared to other functionals from its series, resulting in improved thermochemical estimates including non-covalent interactions.[59] While the B3LYP functional only contains

optimizations for short-range interactions, ωB97-XD is a range separated functional which allows for separate characterization of short range and long range, non-covalent interaction. These two methods should provide insight into the accuracy of B3LYP describing the peptide fragment system.

A summary of the stability gap results from the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) geometry optimizations and subsequent re-optimizations at the M06-2X/6-311++G(d,p) and ωB97-XD/6-311++G(d,p) levels of theory is shown in **Table 1** (A full summary of the results of the B3LYP calculations are available in **Tables 2-5** in Chapter 4 while M06-2X and ωB97-XD re-optimized calculations are available in the Appendix **Tables 10 - 15**). Stability gaps are reported as ΔΔG corresponding to the change in relative 298K corrected free energy.

Table 1 Relative free energy stabilities (kJ/mol) are reported for AK^+, AO^+, and AB^+ as calculated by the B3LYP, M06-2X, and ωB97-XD methods. The results for the AZ^+ have been omitted as they do not have an experimental spectrum available for comparison.

	B3LYP		M06-2X		ωB97-XD	
AK^+ Isomer	**Conformer**	**ΔΔG**	**Conformer**	**ΔΔG**	**Conformer**	**ΔΔG**
Diketopiperazine	diket2	0.0	diket2	0.0	diket2	0.0
Lactam	lac5	3.8	lac5	4.3	lac5	9.3
Oxazolone	oxa2	74.3	oxa2	75.8	oxa2	79.2
AO^+ Isomer						
Lactam	lac5	0.0	lac5	0.0	lac5	0.0
Diketopiperazine	diket2	14.2	diket2	18.4	diket2	15.2
Oxazolone	oxa2	84.1	oxa2	83.1	oxa2	83.7
AB^+ Isomer						
Lactam	lac5	0.0	lac5	0.0	lac2	0.0
Diketopiperazine	diket2	9.9	diket2	10.7	diket2	7.5
Oxazolone	oxa2	100.1	oxa2	96.0	oxa2	102.0

As shown in **Table 1**, all three methods predict the same isomer stability ordering and similar relative stabilities between those isomers. Further, the methods

overwhelmingly agree on the protonation site and hydrogen bonding scheme for the lowest conformer type of each isomer. The three methods predict the lowest lying AK^+ isomer is a diketopiperazine of the **diket2** conformer type. All methods predict a relatively close lactam isomer stability ($\Delta\Delta G$ < 10 kJ/mol) at the **lac5** conformer type. The oxazolone isomer is predicted to be significantly less stable than the diketopiperazine and lactam isomers. All predicted oxazolone isomers are **oxa2** conformers and lie > 74 kJ/mol higher in free energy than the lowest energy diketopiperazine isomer. The hydrogen bonding schemes for each of the lowest lying AK^+ isomers are shown here (See **Figure 11**).

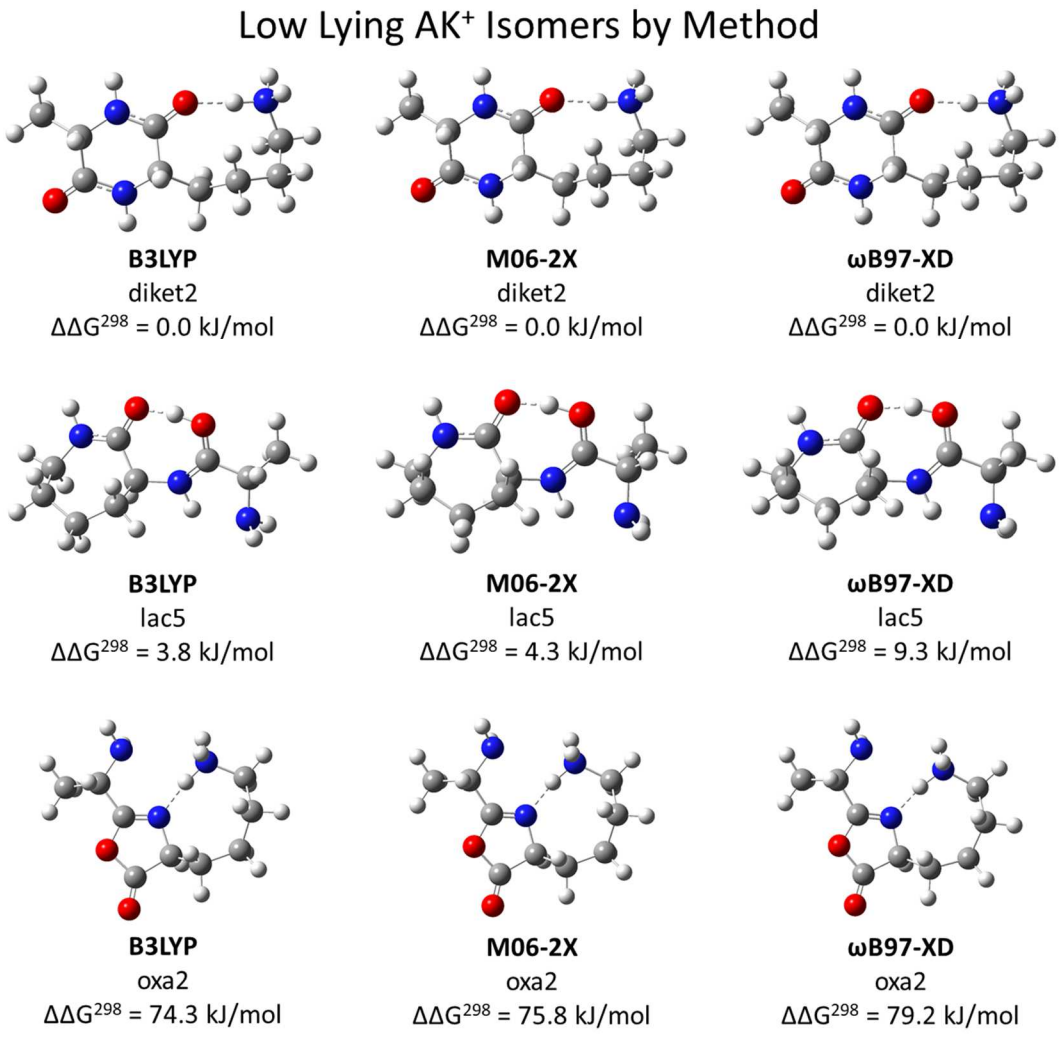


Figure 11 Lowest lying AK^+ structure for each isomer type as computed by the B3LYP (Left), M06-2X (Middle), and ωB97-XD (Right) methods.

The **diket2** conformer has a hydrogen bond between the protonated **K** side chain and the nearby ring carbonyl. The B3LYP structure shows the greatest distance between the donor and acceptor of 1.62 Å. The M06-2X and ωB97-XD methods predict somewhat closer distances of 1.55 and 1.58 Å, respectively. The **lac5** conformer has a clear hydrogen bond between the protonated alanine carbonyl and the lactam ring carbonyl. The hydrogen bonding lengths show similar variation to that seen in the **diket2** structures. The B3LYP calculated structure has a longer coordination distance at 1.42 Å while the M06-2X and ωB97-XD structures predict closer interactions of 1.34 and 1.40 Å, respectively. Some hydrogen bonding interaction is suggested between the amide N-H and the N-terminus, but all three methods predict relatively far coordination distances (> 1.99 Å). The **oxa2** conformer shows a single hydrogen bond interaction between the protonated **K** side chain and the ring nitrogen. Again, the B3LYP method predicts slightly longer range interaction (1.63 Å), but all three methods predicted the distance within 0.03 Å of one another.

The AO^+ stabilities are reversed from AK^+ for the lactam and diketopiperazine isomers. However, the conformer types are all maintained from AK^+ as the lowest energy structures for each isomer (See Appendix, **Figure 39**). The diketopiperazine isomer lies 14.2, 18.4, and 15.2 kJ/mol higher in free energy than the lactam isomer as computed by the B3LYP, M06-2X, and ωB97-XD methods, respectively. The oxazolone isomers are less stable than their AK^+ counterparts. The B3LYP, M06-2X, and ωB97-XD methods predict the oxazolone isomer lies 84.1, 83.1, and 83.7 kJ/mol higher in free energy, respectively.

Interestingly, the AO^+ **diket2** conformer appears to share a nearly identical hydrogen bonding scheme between the three methods (ΔÅ < 0.01) while the **lac5** and

oxa2 conformers display the same hydrogen bond length variation observed in the AK^+ conformers. This indicates that the M06-2X calculation predicts the most stable lactam isomer, and the B3LYP structure has the lowest relative stability. However, the difference in relative stability is only 4.2 kJ/mol and did not result in the B3LYP method incorrectly ordering isomers or misidentifying conformer types.

The AB^+ stabilities follow the same trend as the AO^+ stabilities (See Appendix, **Figure 40**). All three methods identify the lowest energy AB^+ structure as a lactam isomer. The B3LYP, M06-2X, and ωB97-XD methods calculate the diketopiperazine structure lies 9.9, 10.7, and 7.5 kJ/mol higher in free energy than the lactam isomer, respectively. The oxazolone isomer lies 100.1, 96.0, and 102.0 kJ/mol higher in free energy than the lactam isomer as computed by the B3LYP, M06-2X, and ωB97-XD methods. The AB^+ stabilities provide the lone exception to the shared conformer types established in the larger b_2^+ calculations. The B3LYP and M06-2X methods predict the lowest energy lactam is a **lac5** while the ωB97-XD method indicates the structure is a **lac2** conformer (See **Figure 12**). The three methods agree that other low energy structures are **diket2** and **oxa2** conformers.

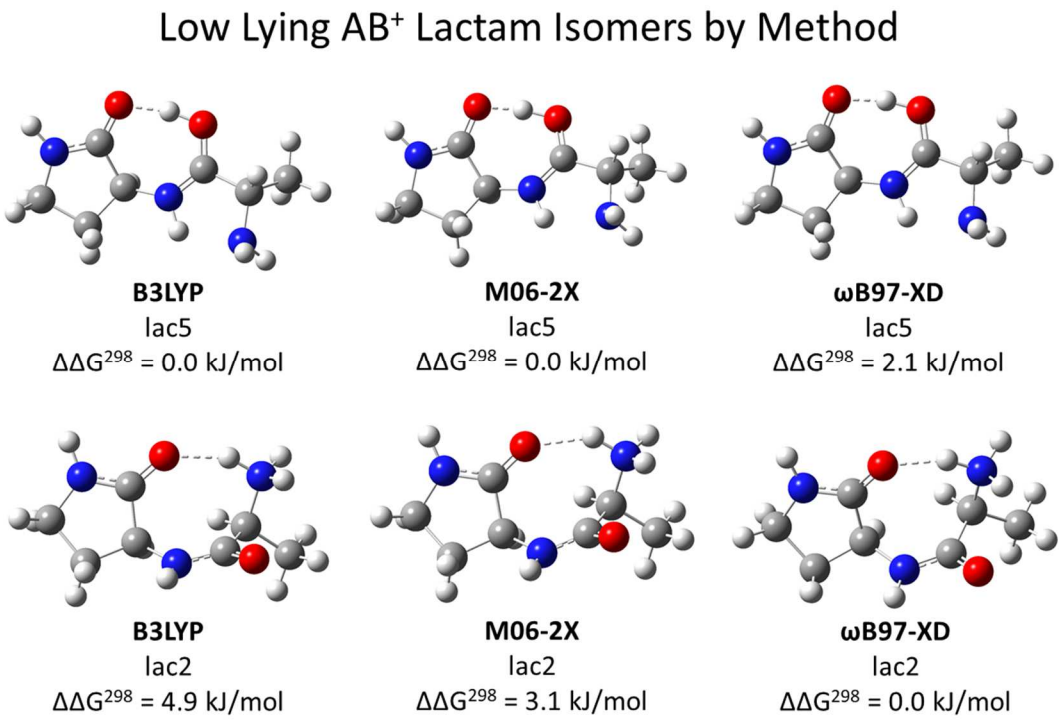


Figure 12 Comparison of the lowest energy conformers of the oxazolone isomer from the B3LYP (Left) M06-2X (Middle), and ωB97-XD (Right) methods.

The ωB97-XD and B3LYP methods both predict relatively small stability gaps between the two lactam isomers in opposite direction, resulting in a net relative energy leveling of ~7 kJ/mol. The M06-2X method optimized a structure preference more similar to the B3LYP optimized structure, but presented a slightly reduced stability gap (3.1 kJ/mol). The B3LYP computed **lac5** structure shows stronger hydrogen bonding interaction between the protonated carbonyl and ring carbonyl (ΔÅ > 0.09) as compared to the ωB97-XD computed structure. However, the two methods produced similar hydrogen bond interaction between the **lac2** computed structures (ΔÅ < 0.02). The difference in treatment of the protonated carbonyl between the two methods could result in the shifted stability observed in the ωB97-XD computed structures.

Further determination of method based calculation differences will draw on their treatment of predicted spectra. A spectral comparison of each low lying AK^+ isomer with

the experimental AK^+ spectrum highlights the differences in predicted absorptions for each method (See **Figure 13**). The difference immediately noticeable in the B3LYP computed spectrum is the treatment of the C=N amide stretch and NH_2 bending and twisting modes. The B3LYP method places these two absorption modes resembling a doublet at 1604 cm^{-1} and 1642 cm^{-1}. The other two spectra indicate these absorption modes will overlap into a single sharp peak of medium-high intensity at 1611 cm^{-1} and 1630 cm^{-1} for the M06-2X and ωB97-XD methods, respectively. Unfortunately, these absorptions will not be instructive in differentiating the three methods' accuracy as comparison to the experimental spectrum reveals no intensity above 1750 cm^{-1}, indicating the experimental fragment is not an oxazolone.

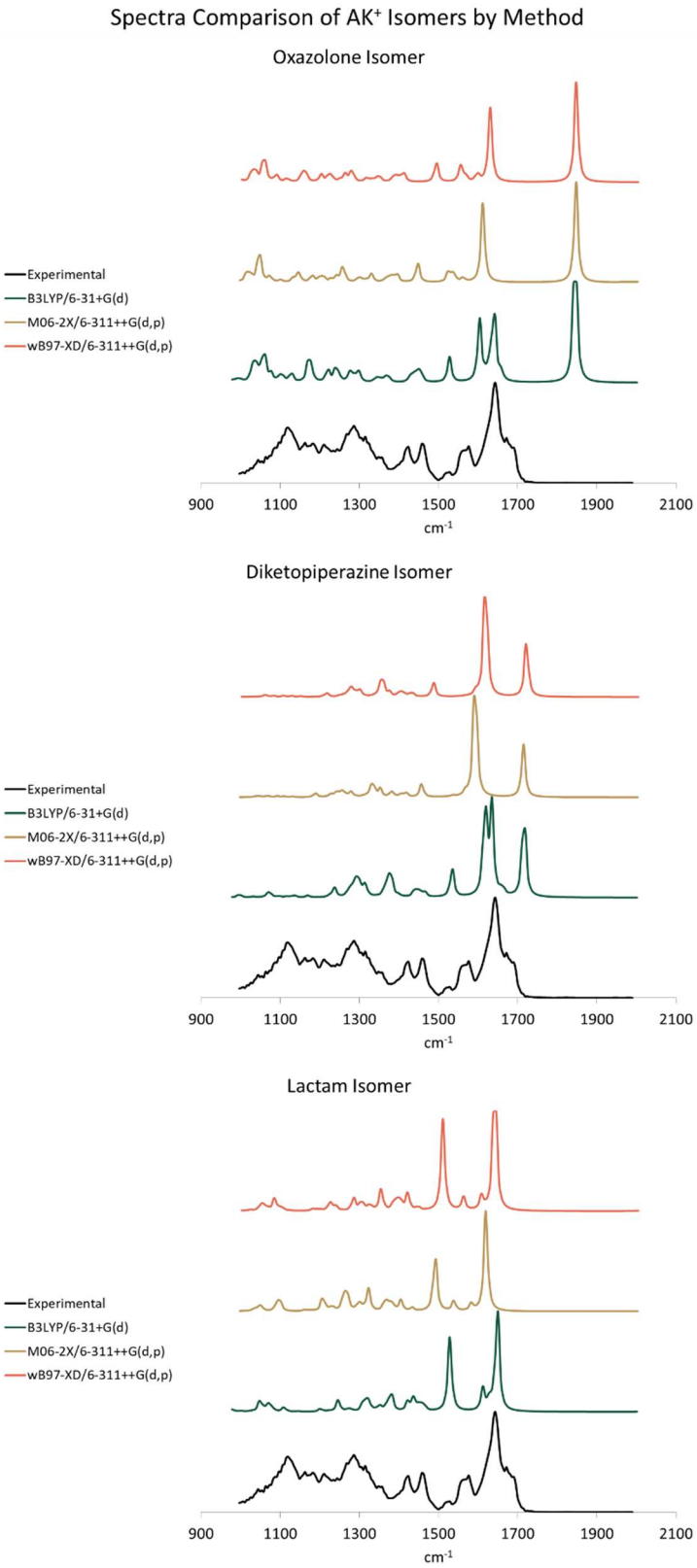


Figure 13 Calculated spectra for the AK^+ Oxazolone (Top), Diketopiperazine (Middle), and Lactam (Bottom) isomers as compared to the experimental spectrum collected from CLIO.

The diketopiperazine spectra reveal a similar story to that of the oxazolone. The B3LYP calculated spectrum predicts that the hydrogen bound carbonyl stretching mode is split into a doublet corresponding to in-plane and out-of-plane NH_3 bending with respect to the carbonyl at 1619 cm^{-1} and 1641 cm^{-1}. This peak is shown to be relatively broad, though, and provides decent overlap with the experimental absorbance at 1640 cm^{-1}. Again, both of the other methods predict a single carbonyl stretch without any sort of splitting, resulting in a sharper peak. The M06-2X calculated spectrum red-shifts this band down to 1585 cm^{-1}. The ωB97-XD computed spectrum places this carbonyl stretch at 1615 cm^{-1}, closer to the lower energy absorption in the B3LYP spectrum, but significantly red-shifted from the experimental band. All three methods predict medium intensity absorptions around 1720 cm^{-1} corresponding to the other carbonyl stretch that is not present in the experimental spectrum.

The lactam spectra show the highest similarity between the three methods. All three methods predict two high intensity, sharp absorptions corresponding to the two carbonyl stretching modes, with a low energy shoulder on the higher energy carbonyl. The placement of these peaks differs between the three methods. The B3LYP spectrum shows the best agreement with the experimental spectrum. The high energy carbonyl stretch absorbs at 1650 cm^{-1} and the lower energy carbonyl absorbs at 1528 cm^{-1}. The ωB97-XD spectrum shows better overlap with the experimental spectrum in the high energy carbonyl (1645 cm^{-1}) but red-shifts the lower energy carbonyl to 1510 cm^{-1}, where the experimental spectrum does not absorb. The M06-2X spectrum performs poorly, predicting carbonyl stretch absorptions at 1590 $cm^{-1}$and 1500 cm^{-1}. The experimental spectrum clearly shows lack of absorbance in these regions.

The spectra for AO^+ and AB^+ confirm similar trends established in the AK^+ spectra (See Appendix, **Figures 41** and **42**). The varying methods did not produce significantly

different spectral results. However, the differences shown—typically red-shifting in carbonyls—resulted in B3LYP generally maintaining the best agreement with the experimental spectrum. The lactam spectrum for AB^+ confirms the accuracy of the B3LYP and M06-2X methods in identifying the **lac5** conformer as the local minimum as opposed to the **lac2** conformer (See **Figure 14**). The ωB97-XD computed spectrum predicts a somewhat broad doublet corresponding to partial overlap of the carbonyl stretching modes at 1743 cm^{-1} and 1720 cm^{-1}, well beyond the absorbing region of the experimental spectrum. The other two methods predict carbonyl absorptions closer to the broad, experimental absorption at 1695 cm^{-1}.

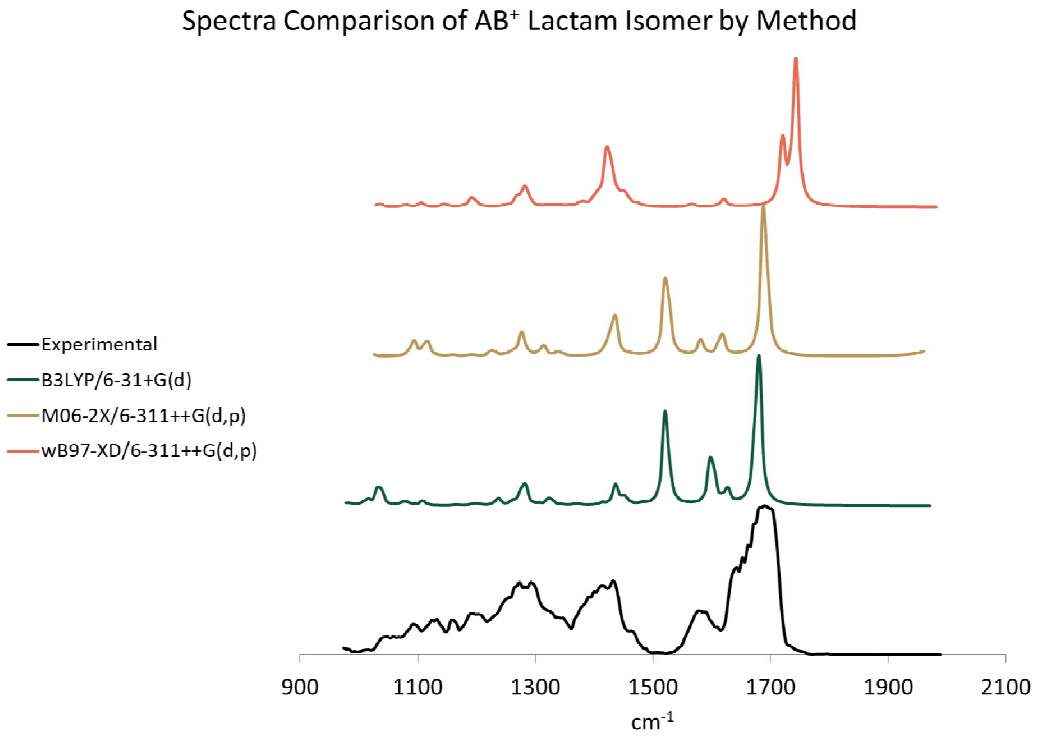


Figure 14 Comparison of the lowest energy AB^+ lactam isomer by method as compared to the experimental spectrum from CLIO.

Conclusion and Future Works

The availability of a NIST recommended scaling factor should not be the sole reason for using one basis over another, the additional descriptiveness should match the additional time investment of the increased robustness. This computational study finds that the addition of p-polarization does not dramatically increase the accuracy of calculations and tends to poorly describe protons shared between two basic sites. IRMPD action spectroscopy often provides broad absorptions that lack the refinement seen in theoretical calculations. Whatever specificity is gained by the addition of p-polarization will be overshadowed by the limit of quantitation for the data it is compared to. Therefore, further re-optimizing to the B3LYP/6-31+G(d,p) level of theory would be a waste of computational resources for this system.

Additional time should be invested into exploring the effects of various basis set parameters on spectral performance. Specifically, the effect of double-ζ versus triple-ζ descriptions for valence atoms and the addition of further diffuse and polarization functions on predicting spectra. This study also limited the type of basis set to Pople basis sets. Other studies have shown the efficacy of correlation-consistent and polarization-consistent basis sets on the prediction of vibrational frequencies for small biomolecules.[62] In the interim, B3LYP/6-31+G(d) computed spectra provide reasonable accuracy with relatively inexpensive computing time.

The three methods studied did not provide significantly different results. The three methods predicted the correct conformer in almost, and when ωB97-XD predicted the wrong conformer, the correct one was very close in relative stability (2.1 kJ/mol). Spectral predictions also remained largely inconclusive. There was certainly absorption wavelength shifting between the three methods. However, as discussed in the basis set comparison, the scaling factor can compensate for these differing wavelengths. Though the

absorptions shifted relative to the other methods, they did not shift relative to other peaks in the same spectrum meaning a scaling factor will adjust the absorptions higher or lower in energy by the same amount.

Though the 6-311++G(d,p) basis set was necessary for correct energy leveling using the B3LYP method, the 6-31+G(d) method provided similar spectra to the other methods using larger basis sets. A smaller basis set reduces the computation time significantly. The M06-2X and ωB97-XD methods took between 4 and 5 hours per calculation, while the corresponding B3LYP calculation took between 3 and 4 hours. Therefore, without any significant improvement in energetics or spectral prediction, we will continue with the B3LYP method.

Chapter 4: b_2^+ Fragment Ions

Methods

Theoretical Procedures Calculations for the lactam, diketopiperazine, and oxazolone forms of AK^+, AO^+, AB^+, and AZ^+ were computed using the Gaussian 09 suite. Conformer sets were generated for several protonation sites (see **Figure 6**) using the GMMX searching routine in PCModel 9 which varies bond lengths and angles, saving structures with energies within 40 kJ/mol of the minimum energy structure. These conformer sets were imported into Gaussian and used as starting structures for a series of molecular orbital and density functional theory calculations. Final geometries, zero-point energies, and thermal corrections to enthalpy and free energies were computed at the B3LYP/6-31+G(d) level of theory for all species.

The recommended scaling factors for B3LYP frequencies are 0.98 for ZPE and 0.99 for thermal corrections to enthalpy and free energy. Due to the small differences between the recommended scaling factors and the unscaled harmonic frequencies, we chose to use the unscaled harmonic frequencies to derive ZPE and thermal corrections. The reported energy stabilities were derived by extracting the total electronic energy from a ZPE computation at the B3LYP/6-311++G(d,p) level of theory and converting to the 298K enthalpies and free energies using the B3LYP/6-31+G(d) computed thermal corrections.

The B3LYP/6-31+G(d) level of theory was used in extracting theoretical vibrational spectra for comparison to the experimental IRMPD data. This method and basis set combination does not have a NIST recommended scaling factor. By matching diagnostic peaks to in the theoretical spectrum to their experimental counterparts, a scaling factor of 0.955 was reached. Stick spectra were broadened by a 20 cm^{-1} Gaussian function.

Experimental Procedures IRMPD spectra for the b_2^+ fragment ion species were obtained using the FEL facility, CLIO, in Orsay, France and the FELIX facility in Nijmegen, the Netherlands. Infrared spectra for AK^+ and AB^+ were obtained in the fingerprint region (~1000 cm^{-1} – 2000 cm^{-1}) using the CLIO FEL in a modified Bruker Esquire quadrupole ion trap mass spectrometer set up described elsewhere.[63] AZ^+ fragment ions could not be isolated during the allotted laser time. AK^+ (*m/z* = 200 amu) fragments to peaks at *m/z* = 183 amu (loss of NH_3), 155 (likely the loss of NH_3CO), and 128 (loss of $C_2H_5COCH_3$). AB^+ (*m/z* = 172 amu) fragments to 127 (loss of NH_3CO) and 100 (loss of $C_2H_5COCH_3$).

Fragment spectra were obtained as IRMPD action spectra. Dilute solutions (~10 μM) of the corresponding tetrapeptide were prepared in acidified (~1% Formic Acid) 50:50 MeOH:H_2O (CLIO) or 50:50 C_2H_3N:H_2O (FELIX) and injected into the mass spectrometer via the electrospray ionization source. Mass spectrometer conditions and ion focusing optics are tuned to optimize the corresponding tetrapeptide. Precursor tetrapeptides are mass selected and allowed to undergo CID fragmentation to yield the corresponding b_2^+ ion. Fragment ions are then mass selected and irradiated with the infrared FEL. The IRMPD fragment intensities are monitored as wavelengths are scanned. An infrared spectrum is generated by plotting the sum of all fragment ions normalized to the sum of the fragments and the precursor ion (∑ (fragments) / ∑ (fragments + precursor)).

FEL wavelength calibration occurs by splitting the beam and passing a portion through a polystyrene film as a control spectrum in real time. This experimental spectrum is then linearly fit to a known polystyrene spectrum. This linear fit is applied to the analyte as a wavelength correction. The power (mJ) measured from the polystyrene spectrum is fit to the wavelength as a polynomial function. This fit is used to generate power readings for each wavelength scanned. The measured intensity is then divided by the normalized

power to produce the corrected absorbance. Final analyte spectra are then normalized and have a three-point average smoothing factor applied.

Similar procedures were used at the FELIX Facility in Njimegen, Netherlands to obtain spectra for AO^+ and the novel **K** and **O** lactam b_2^+ structures (AK-lac$^+$ and AO-lac$^+$, respectively). AO^+, AK-lac$^+$, AO-lac$^+$ spectra were obtained in the fingerprint region using the FEL in a modified Bruker amaZon ion trap mass spectrometer.[64] The FELIX FEL operates at a 1 GHz macropulse with 10 Hz micropulses. Optical sensors detect the firing laser and open shutters to the ion trap chamber for 200 milliseconds, irradiating trapped ions with 2 pulses of the laser. The laser moves in steps of ~4 cm^{-1}, and six averages are taken at each measured wavelength. FELIX operates under a similar workflow to that of CLIO but required much less concentrated samples (~1 μM). AO^+ (*m/z* = 186 amu) fragments to yield peaks at *m/z* = 141 amu (cleavage at the alanine α-carbon yielding loss of $H_3CCH\text{-}NH_2$), 115 (cleavage at the peptide bond, loss of $NH_2CHCH_3C{=}O$), and 71 (likely the complementary fragment to the previous loss). AK-lac$^+$ (*m/z* = 200 amu) shows fragment intensities at *m/z* = 128 amu (loss of $NH_2CHCH_3C{=}O$) and 84 (loss of $NH_2CHCH_3C{=}ONHCHO$). AO-lac$^+$ (*m/z* = 186 amu) fragments to give peaks at *m/z* = 169 amu (loss of NH_3), *m/z* = 141 amu (loss of CH_3CHNH_2), 115 (loss of $NH_2CHCH_3C{=}O$), and 71 (loss of $C_5H_8NO\text{-}NH_2$).

Wavelength validation occurs online via a grating spectrometer, and an additional correction takes place during data work up from power and wavelength calibrations taken between runs. The difference in expected and reported wavelengths is fit to the expected wavelength as a polynomial function such that no point varies from the fit by more than 0.02 μm, the limit of detection for the undulator. This polynomial fit is applied to the wavelengths scanned, and then converted to wavenumbers for use in captured spectra. A correction for frequency dependent laser power variation is also applied during data

workup. This power correction is performed in the same manner as that applied to the CLIO data.

Materials Alanine Wang resin was purchased from ChemPep Inc (Wellington, FL). FMOC-Alanine-OH, FMOC-Lysine(BOC)-OH, FMOC-Ornithine(BOC)-OH, FMOC-diaminobutanoic acid(BOC)-OH, and FMOC-diaminopropionoic acid(BOC)-OH were purchased from MilliporeSigma (Burlington, MA). L-α-amino-ε-caprolactam and (S)-3-amino-2-piperidone were purchased from Sigma-Aldrich (Burlington, MA) and used without further purification. All materials were used without purification. Tetrapeptides were synthesized with lysine homologs in the second position using a solid state peptide synthesis protocol described elsewhere.[65] Novel AK-lac$^+$ and AO-lac$^+$ structures were synthesized by the Scheerer group (College of William & Mary, Williamsburg, VA) using the purchased lactam and FMOC-Alanine-OH.

Results and Discussion

A comparison of the experimental data from CLIO and FELIX reveal similar spectral features across the captured AK$^+$ and AB$^+$ data and present their differences from the AO$^+$ spectrum (See **Figure 15**). The most notable feature of all three spectra is the striking lack of signal above 1750 cm^{-1}. Without absorbance in the high energy range of the experimental window, the oxazolone structure can be eliminated as a possible observed isomer. The experimental spectra all present similar high intensity, broad absorbance between 1650 cm^{-1} and 1700 cm^{-1}. However, the AO$^+$ spectrum shows a second broad absorbance near 1530 cm^{-1} where AK$^+$ and AB$^+$ lack intensity at this wavelength. Further determination of experimental populations required comparison to theoretical spectra for each of the IRMPD spectra.

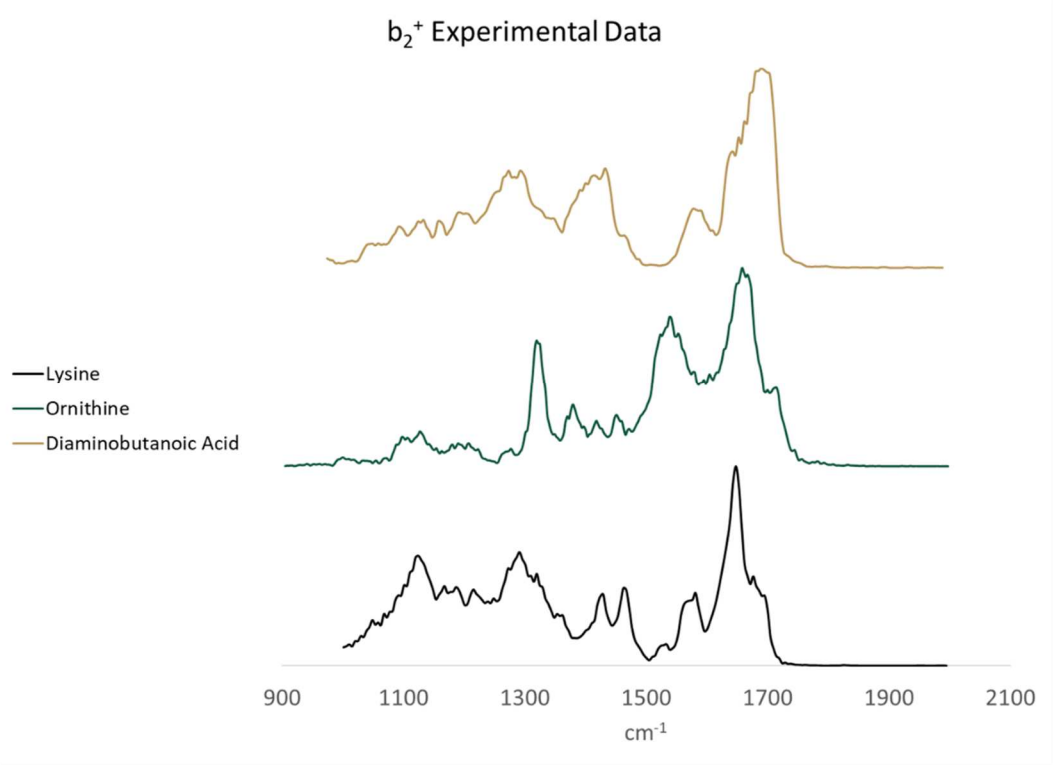


Figure 15 A comparison of the three collected IRMPD spectra from CLIO (AK^+ and AB^+) and FELIX (AO^+) shows AK^+ (Black), AO^+ (Green), and AB^+ (Gold). AZ^+ could not be isolated during the allotted laser time.

Lysine As described earlier, the lowest energy AK^+ isomer is a diketopiperazine of the **diket2** conformer type. The lowest energy lactam isomer lies 3.8 kJ/mol higher in free energy, and the lowest energy oxazolone is 74.3 kJ/mol above the global minimum. The lowest lying conformer type from each isomer is tabulated here (See **Table 2**), and selections from the full conformer search are available in the Appendix (See **Table 16**).

Table 2 The lowest energy conformer of each isomer of AK^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Conformer	ΔG	ΔΔG
ak_b2_diket_077	diket2	-668.2778117	0.0
ak_lac4_024	lac5	-668.2763483	3.8
ak_lac2_003	lac2	-668.2743106	9.2
ak_lac3_002	lac3	-668.2536419	63.5
ak_b2_diket3_021	diket1	-668.2533345	64.3
ak_oxa3_017	oxa2	-668.2495116	74.3
ak_b2_lac_001	lac1	-668.2491537	75.2
ak_lac3_122	lac4	-668.2471757	80.4
ak_b2_diket2_107	diket3	-668.2448444	86.6
ak_b2_oxa_056	oxa1	-668.2397296	100.0
ak_oxa3_069	oxa3	-668.2394658	100.7

Structural evaluation and comparison for the lactam was treated differently to that of the diketopiperazine and oxazolone. The lactam structures should present similar hydrogen bonding schemes across the same conformer types of the different residues studied. The differences present in relative stabilities of these structures are driven by ring size as opposed to favorable non-covalent interactions. Conversely, the diketopiperazine and oxazolone structures maintain their ring size regardless of residue. As such, the driving force in changes to their stability is instead favorable non-covalent interactions available to the amino residue studied.

As expected, each conformer for AK b_2^+ type displays a unique hydrogen bonding scheme (See **Figure 16**). The **diket2**, **lac5**, and **oxa2** are the lowest energy conformers of their respective isomer type and were detailed in the computational comparison. Interestingly, hydrogen bonding schemes are not maintained when the proton shifts from the hydrogen bond donor to the acceptor. As shown in the **diket2** structure, the protonated **K** side chain shows hydrogen bonding character with the nearest carbonyl. If the hydrogen were to move from the **K** side chain to the carbonyl oxygen, the same hydrogen bonding scheme would be maintained. However, the hydrogen bonding scheme shown in **diket1**

indicates that the neutral **K** side chain forms a more favorable interaction with the adjacent ring N-H hydrogen bond donor. This scheme reduces the size of the **diket2** “8-membered ring” to a more favorable “7-membered ring.” This indicates that while the **K** side chain is the optimal site for protonation, an “8-membered ring” is not the ideal hydrogen bonding formation were some other site available.

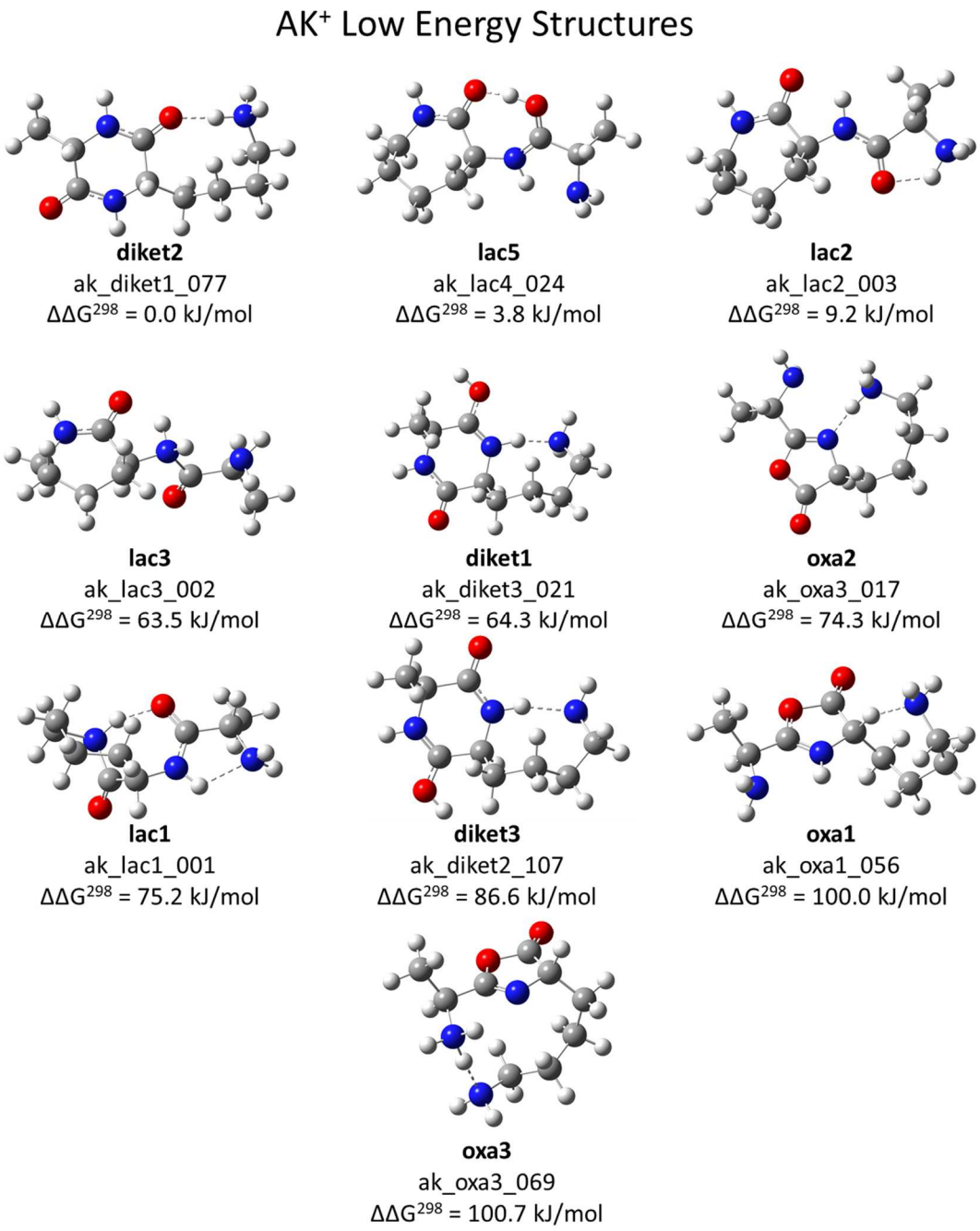


Figure 16 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AK^+.

The spectral comparison of the lowest energy calculated spectra for each isomer and the experimental IRMPD spectrum is shown below (See **Figure 17**). Energetically, the oxazolone should not be present in the experimental spectrum. The oxazolone is over 70 kJ/mol less stable than the diketopiperazine and oxazolone forms and is quickly

eliminated by the lack of signal above 1750 cm^{-1}. The diketopiperazine isomer should be the favored structure. The overlapping amide and carbonyl stretches near 1640 cm^{-1} provide decent overlap with the most intense band of the experimental spectrum. However, the higher energy carbonyl near 1710 cm^{-1} as well as the NH_3 bend at 1535 cm^{-1} are not seen in the experimental spectrum. The lactam isomer (+3.8 kJ/mol) is also an energetically viable structure. The intense carbonyl stretch near 1650 cm^{-1} matches well with the experimental absorption, and the lower energy amide absorption matches the lower energy tail from the experimental spectrum. The lower energy carbonyl stretch, near 1530 cm^{-1} is seen at low intensity in the experimental spectrum. However, the lactam theoretical spectrum fails to predict the medium intensity band observed near 1570 cm^{-1} nor the high energy shoulder on the high intensity peak in the experimental spectrum.

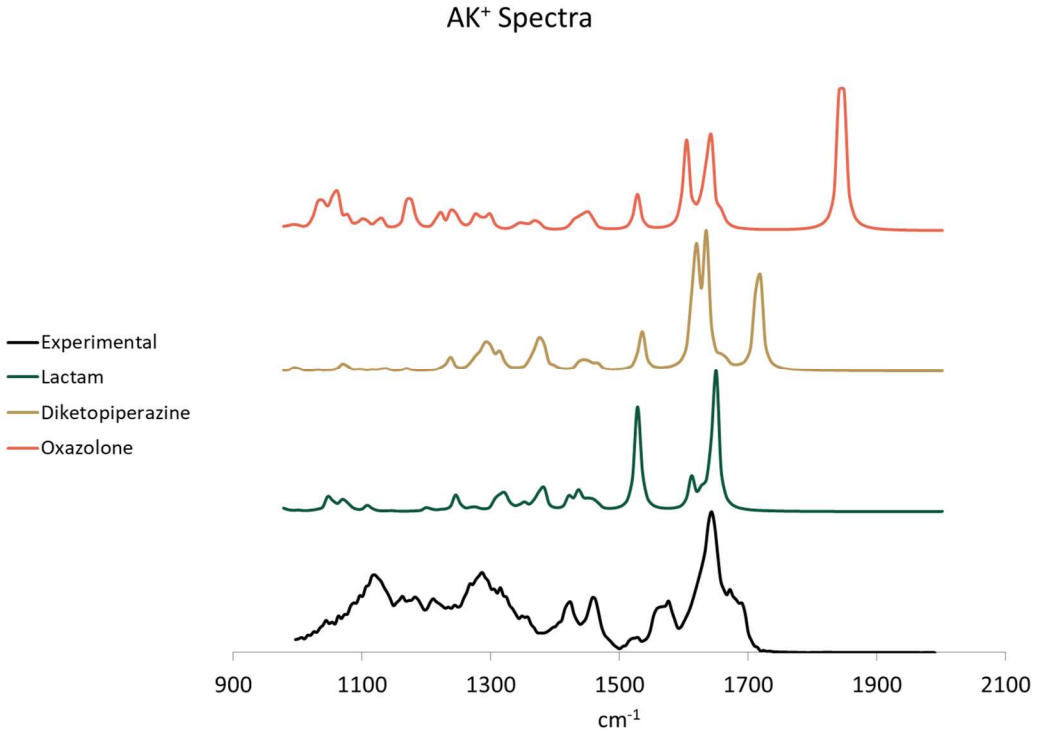


Figure 17 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AK^+.

To resolve the uncertain results from the theoretical spectra, an authentic AK^+ lactam was synthesized for comparison. The spectrum initially taken suffered from broadening issues in the higher energy region. When 3 dB of attenuation was applied the spectrum gained clarity. Additional attenuation was added but resulted in increased broadening from depression of signal. As such the spectrum remains somewhat broad but remains the best clarity available under these lasing conditions. The theoretical lactam spectrum is compared to the experimental spectrum for validation (See **Figure 18**).

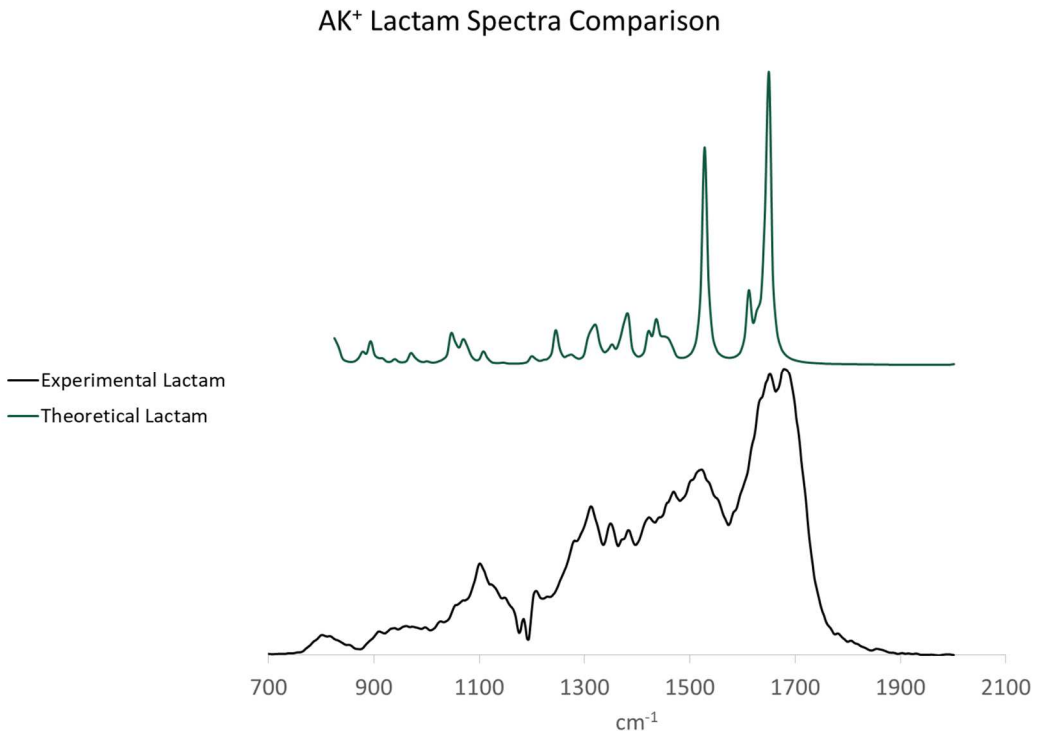


Figure 18 Comparison of the experimental AK^+ lactam spectrum collected at FELIX (Black) and the theoretical lactam spectrum (Green).

The theoretical spectrum predicts well the location of the intense peaks from the experimental spectrum. The carbonyl stretch appears slightly red-shifted from the expected location in the experimental spectrum. This is likely an artifact from the chosen scaling factor. The scaling factor was selected based on data from the CLIO facility. CLIO applies a linear wavelength correction that can skew the calibration, especially in the upper

fingerprint region. The second identifiable peak in the experimental spectrum is likely the amide stretch shown in the theoretical spectrum. Due to broadening, characterization of the lower energy peaks is difficult, though the small intensity absorbers do share a similar loss of absorbance near 1200 cm^{-1}.

The experimental lactam spectrum is also useful for isomer determination of the fragment spectrum. An overlay of the novel lactam and fragment spectra reveal the extent of shared features (See **Figure 19**). The intense absorbance in the fragment spectra overlays nicely with the lower energy doublet peak from the lactam spectrum. Additionally, a small shoulder on the high energy side of the peak near 1300 cm^{-1} in the fragment spectrum aligns nicely with a corresponding intensity in the lactam spectrum. The lactam spectrum has reduced absorbance at the peak near 1575 cm^{-1}. Similarly, the intense absorbance at 1530 cm^{-1} in the lactam spectrum is reduced to a significantly less intensity peak in the fragment spectrum.

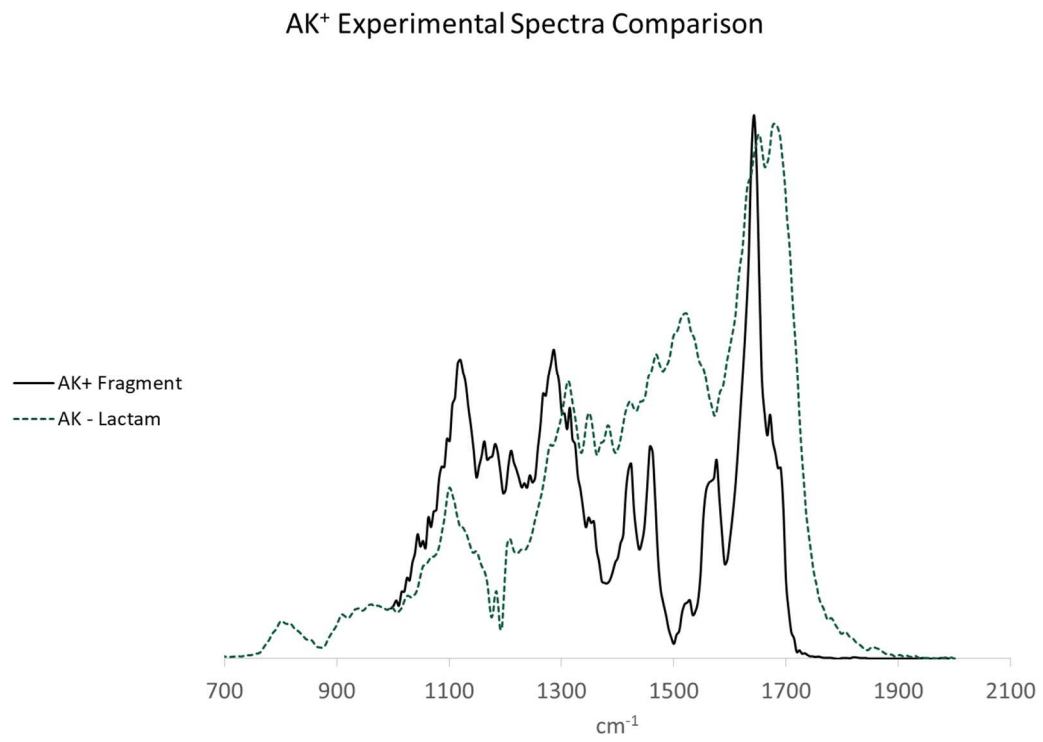


Figure 19 An overlay of spectra from the b_2^+ fragment at CLIO (Solid Black) and the novel lactam synthesis at FELIX (Dashed Green) for lysine.

Taken together, this likely indicates the measured fragment spectrum exists as a mixture, containing both the lactam form and another isomer. The high intensity peak from the lactam structure likely forms the high energy shoulder of the carbonyl stretch in the fragment spectrum. The same is true of the amide stretch, with lower energy absorptions contributing to the broadening in the 1100 cm^{-1} – 1300 cm^{-1} region of the fragment spectrum. The lactam isomer likely exists as a small part of the mixture. This is predicted energetically by the small but clear preference (3.8 kJ/mol) of the diketopiperazine.

To test this, combined spectra were developed from the lowest energy diketopiperazine and lactam isomers. The spectra were combined by varying amounts and then compared to the fragment spectrum for accuracy. These combined spectra are shown below (See **Figure 20**). The combined spectra containing 90% and 65% diketopiperazine mimic the intense absorbing mode in the fragment spectrum well.

However, these spectra continue to omit the fragment peak at 1580 cm^{-1} and include an additional carbonyl stretch at 1720 cm^{-1} that is entirely absent in the experimental spectrum. Spectral overlays of the diketopiperazine and lactam isomers with the experimental data reveal distinct peaks in each that are not present in the experimental spectrum (See Appendix, **Figures 43** and **44**). Additional measurements will need to be taken for exact confirmation of the AK^+ fragment.

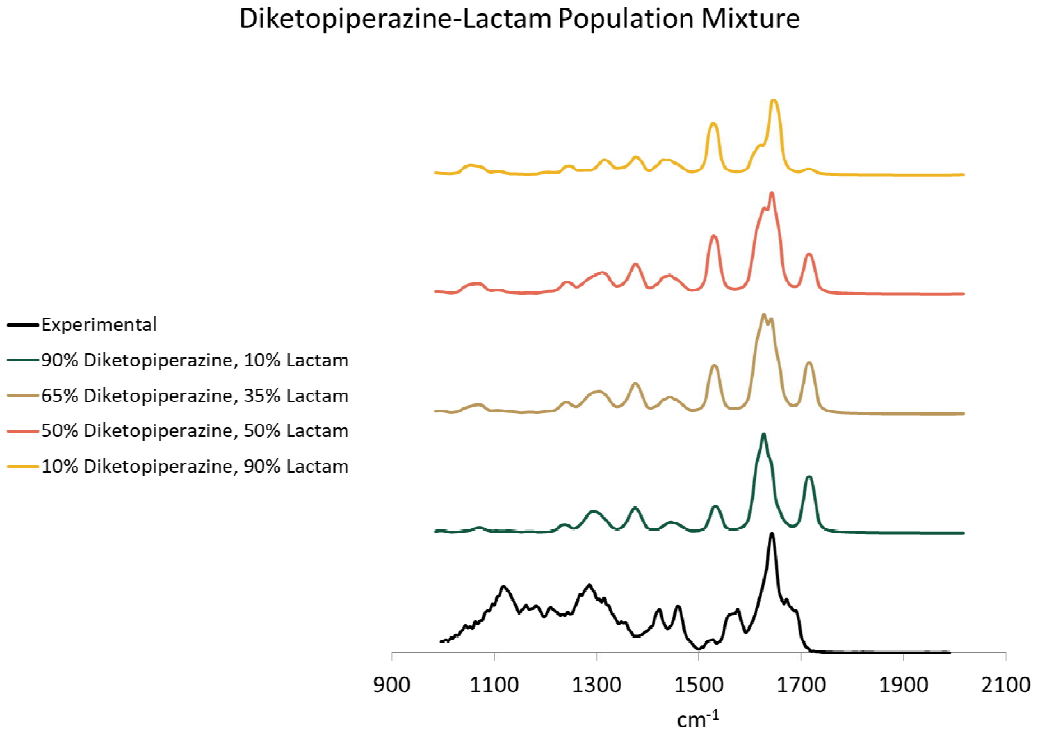


Figure 20 Mixed diketopiperazine-lactam populations are evaluated against the AK^+ experimental IRMPD spectrum (Black). The mixtures evaluated are: 90% diketopiperazine, 10% lactam (Green); 65% diketopiperazine, 35% lactam (Gold); 50% diketopiperazine, 50% lactam (Red); and 10% diketopiperazine, 90% lactam (Yellow).

Ornithine AO^+ shares similar stabilities to those observed from AK^+. The lowest energy isomer is a lactam of the **lac5** conformer type. The lowest energy diketopiperazine and oxazolone isomers were located 14.1 and 84.1 kJ/mol higher in free energy than the lactam global minimum. The energies of the lowest lying conformer types are presented

here (See **Table 3**) and selections from the full conformational search are available in the Appendix (See **Table 17**).

Table 3 The lowest energy conformer of each isomer of AO^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Conformer	ΔG	ΔΔG
ao_lac4_012	lac5	-628.9839413	0.0
ao_lac2_001	lac2	-628.9806749	8.6
ao_diket2_010	diket2	-628.9785503	14.2
ao_diket1_305	diket4	-628.9606707	61.1
ao_diket1_139	diket1	-628.9586983	66.3
ao_lac3_002	lac3	-628.9551428	75.6
ao_lac1_005	lac1	-628.9549317	76.2
ao_oxa2_001	oxa2	-628.9518923	84.1
ao_diket3_339	diket3	-628.9482096	93.8
ao_oxa1_165	oxa1	-628.9447333	102.9
ao_diket3_360	bicyclic1	-628.9412895	112.0
ao_oxa3_069	oxa3	-628.9349033	128.7
ao_oxa1_052	bicyclic2	-628.9342257	130.5
ao_diket1_286	bicyclic3	-628.932969	133.8

As predicted by the ornithine effect, the most stable isomer of AO^+ is the lactam. Both the **lac5** and **lac2** (8.6 kJ/mol) are computed to have a higher stability than the diketopiperazine isomer. This stability change shows that the 6-membered lactam ring creates a more stable structure than the potential hydrogen bonding motifs available to the shorter **O** side chain. The optimized structures share hydrogen bonding schemes with the AK^+ structures of the same conformer type (See **Figure 21**). The AO^+ conformer search identified a new hydrogen bonding scheme not seen in the AK^+ search. Like AK^+, the lowest energy diketopiperazine is a **diket2** conformer protonated on the **O** side chain. The **diket2** structure has the same hydrogen bonding scheme as AK^+. However, another similar structure was located from the **diket1** starting conformer. In this structure, the amide proton adjacent to the **O** side chain shifted to the side chain amino group creating a **diket4** conformer type. This appears 4.2 kJ/mol lower in free energy the **diket1** structure

with the same hydrogen bonding scheme. Additional bicyclic structures were found in the high energy range. These conformer types were not investigated further as they only existed at sufficiently extended free energies.

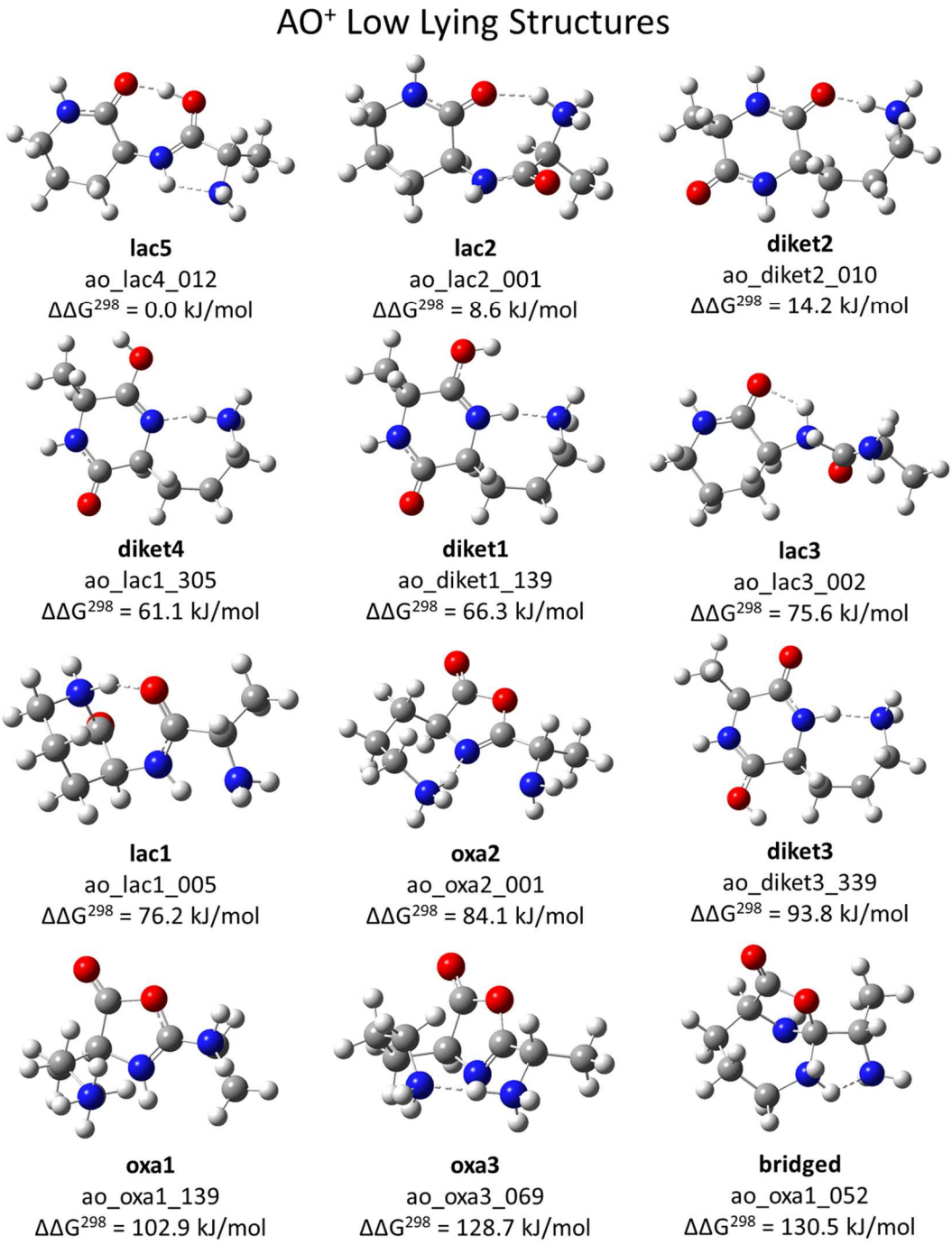


Figure 21 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AO^+.

The spectra produced by these isomers yield more definitive results as compared to the experimental spectrum (See **Figure 22**). The experimental spectrum suffers from a similar broadening problem shown in the AK^+ lactam spectrum. AO^+ retained signal with higher attenuation though, giving a sufficiently clear spectrum at 5 dB. Despite the increased attenuation, the higher energy region remains relatively broad. Three peaks are clearly discernable in the experimental spectrum, which are characterized by three intense absorbing regions at 1670 cm^{-1}, 1545 cm^{-1}, and 1325 cm^{-1}. The first peak, likely a carbonyl, also shows a clear shoulder of medium intensity extending at 1715 cm^{-1}. No signal is detected above this shoulder.

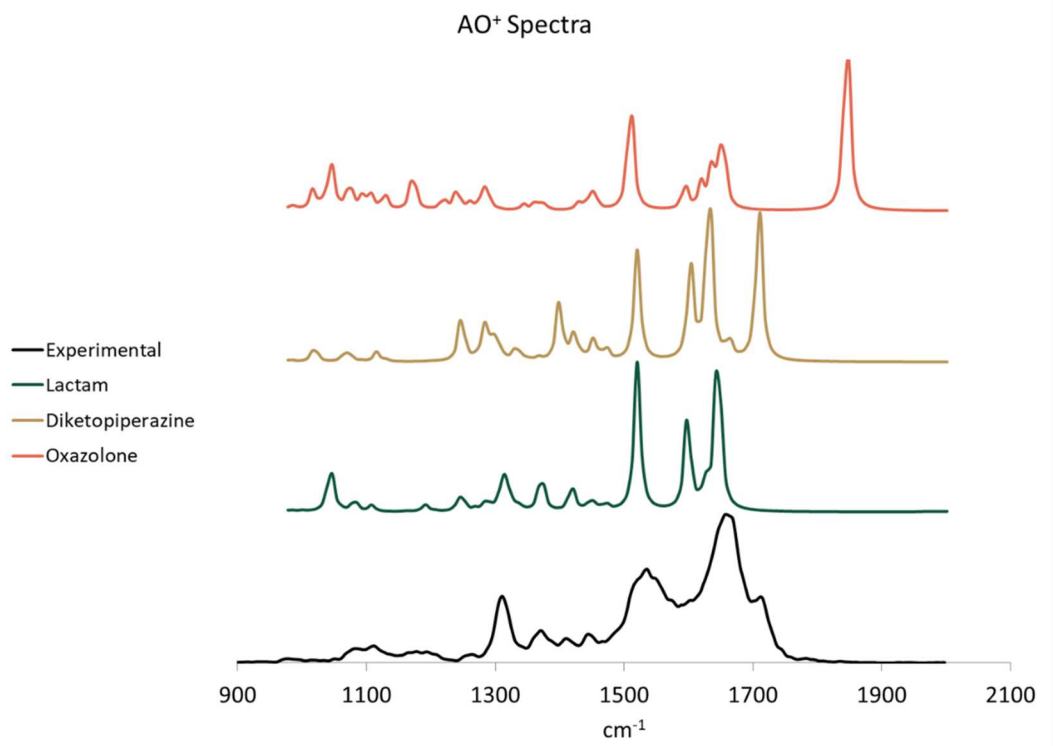


Figure 22 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AK^+.

The lack of a blue-shifted carbonyl immediately disqualifies the oxazolone as a potential observed isomer. The diketopiperazine and lactam spectra contain many similar features. They both show a strong amide stretch absorbance than aligns with the 1545

cm^{-1} absorption in the experimental spectrum. Both also show a second amide absorption near 1600 cm^{-1} nearby a more intense carbonyl absorption near 1640 cm^{-1}. The lactam predicts no other higher energy absorption while the diketopiperazine spectrum shows a second strong carbonyl peak at 1720 cm^{-1}. This peak could be the shoulder seen in the experimental spectrum, though calculated spectra typically predict the carbonyl stretches with accuracy and the shoulder appears at a significantly reduced intensity.

As both low energy spectra predict similar absorptions in line with the fragment spectrum, an authentic AO^+ lactam was synthesized for spectral comparison. The authentic lactam spectrum was taken under the same laser calibration and attenuation settings as the fragment spectrum. An overlay of the two spectra confirms the lactam isomer is the sole AO^+ structure formed during fragmentation (See **Figure 23**). The two experimental spectra are nearly identical. There is some intensity discrepancy in the lower energy region. FELs naturally produce more power at lower wavenumbers. As these spectra were run sequentially, additional calibration was not taken between runs which could account for the intensity differences.

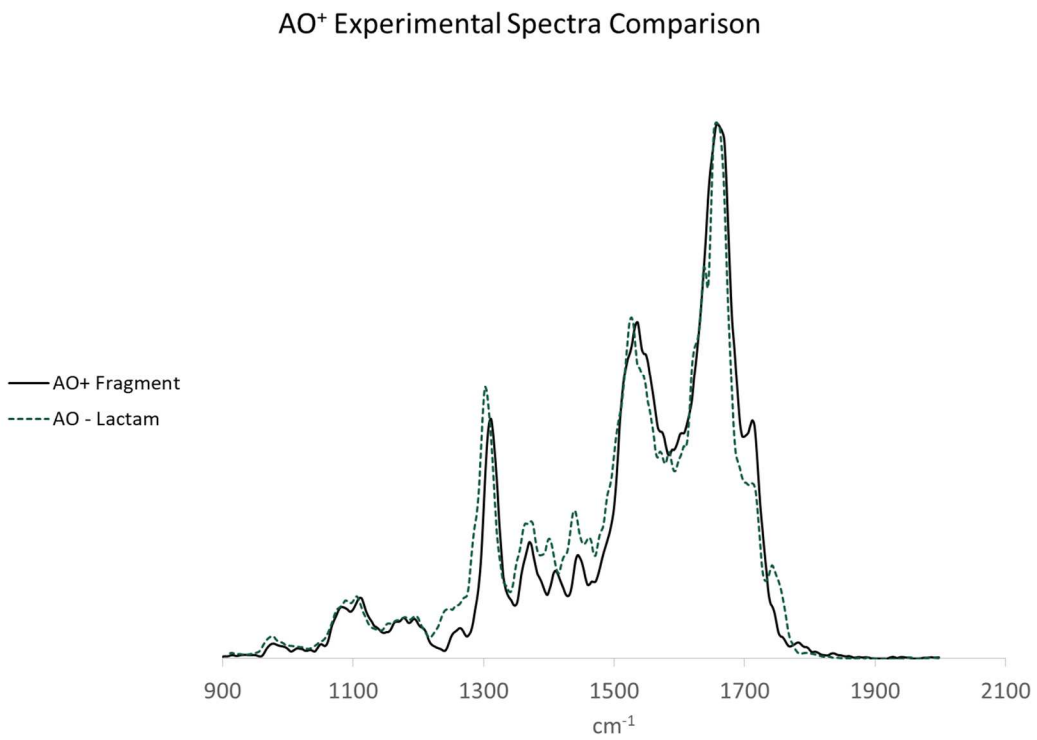


Figure 23 An overlay of spectra from the b_2^+ fragment at CLIO (Solid Black) and the novel lactam synthesis at FELIX (Dashed Green) for AK^+.

Diaminobutanoic Acid As noted in the computational comparison, the lowest energy isomer of AB^+ is a lactam. The diketopiperazine and oxazolone isomers are located 9.9 kJ/mol and 100.1 kJ/mol higher in free energy. Relative stabilities of the lowest energy structures for each conformer type are tabulated here (See **Table 4**) and selected excerpts from the full conformational search are available in the Appendix (See **Table 18**).

Table 4 The lowest energy conformer of each isomer of AB^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Conformer	ΔG	ΔΔG
ab_lac1_001	lac4	-589.6818299	0.0
ab_lac2_007	lac2	-589.679969	4.9
ab_diket3_009	diket2	-589.6780628	9.9
ab_diket2_001	diket2	-589.6780594	9.9
ab_diket1_071	diket4	-589.65817	62.1
ab_diket1_084	diket1	-589.6534834	74.4
ab_lac3_007	lac3	-589.6495093	84.9
ab_oxa1_004	lac1	-589.6482797	88.1
ab_diket3_130	diket3	-589.647566	90.0
ab_oxa3_003	oxa2	-589.6437169	100.1
ab_oxa1_030	oxa1	-589.6422297	104.0
ab_oxa3_005	oxa3	-589.6311581	133.0

Final structures of each low-lying conformer show the same hydrogen bonding schemes discussed (See **Figure 24**). The calculated AB^+ structures follow the same conformer ordering as the AO^+ isomers except in the oxazolone stability. The **oxa2** conformer type is less stable (~10 kJ/mol) than its AO^+ counterpart. This shift is likely due to the decreased stability of the 5-membered hydrogen bonding scheme as compared to the 6-membered scheme in AO^+. This decrease in stability extends to the lactam structures as well. The AB^+ **lac5** structure is ~5 kJ/mol less stable than the AO^+ **lac5**. Again, this change comes from the reduction in the lactam ring size due to the loss of a methylene in the **B** side chain. The relative stability differences may be amplified by increased stability of the **diket2** conformer type. The AB^+ **diket2** adopts the same hydrogen bonding scheme its longer homologs, but the **B** side chain forms a 6-membered hydrogen bond as opposed to the 7- and 8-membered schemes preferred by **O** and **K**.

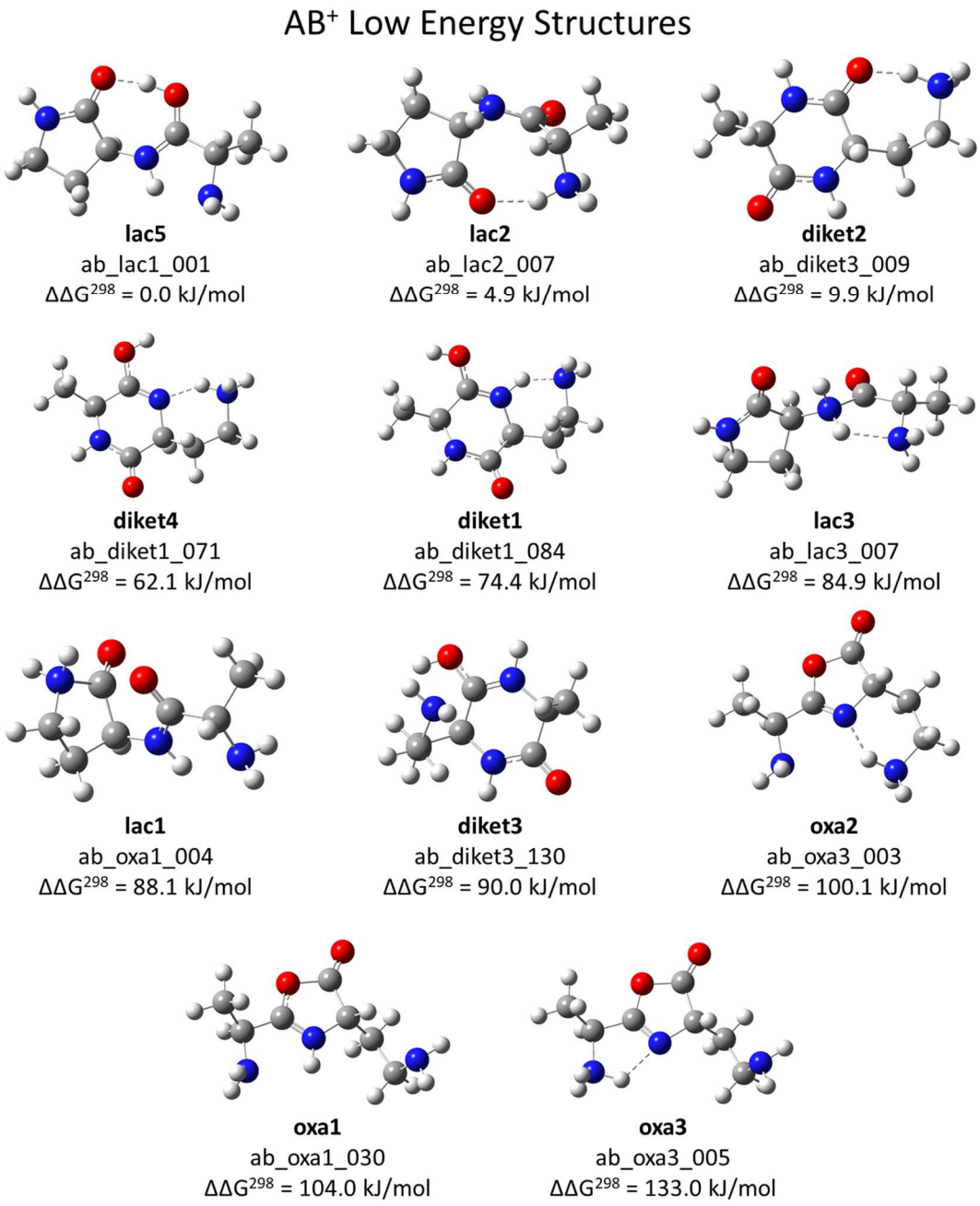


Figure 24 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AB^+.

As in the AK^+ spectra, none of the low energy conformers replicate the absorption bands detected in the AB^+ spectrum (See **Figure 25**). The lactam isomer provides the best coverage, showing intense absorption from the carbonyl at 1680 cm^{-1} with another sharp, low intensity peak at 1595 cm^{-1}. These predict the two higher energy absorptions

seen in the AB^+ spectrum. As with the other **lac5** spectra, there is a sharp peak of medium intensity from the alanine carbonyl stretch near 1520 cm^{-1} that is absent in the experimental spectrum. The diketopiperazine does a poorer job predicting the experimental data. The carbonyl stretches appear as a wide doublet at 1640 cm^{-1} and 1720 cm^{-1}, predicting no absorption where the AB^+ spectrum shows its highest intensity.

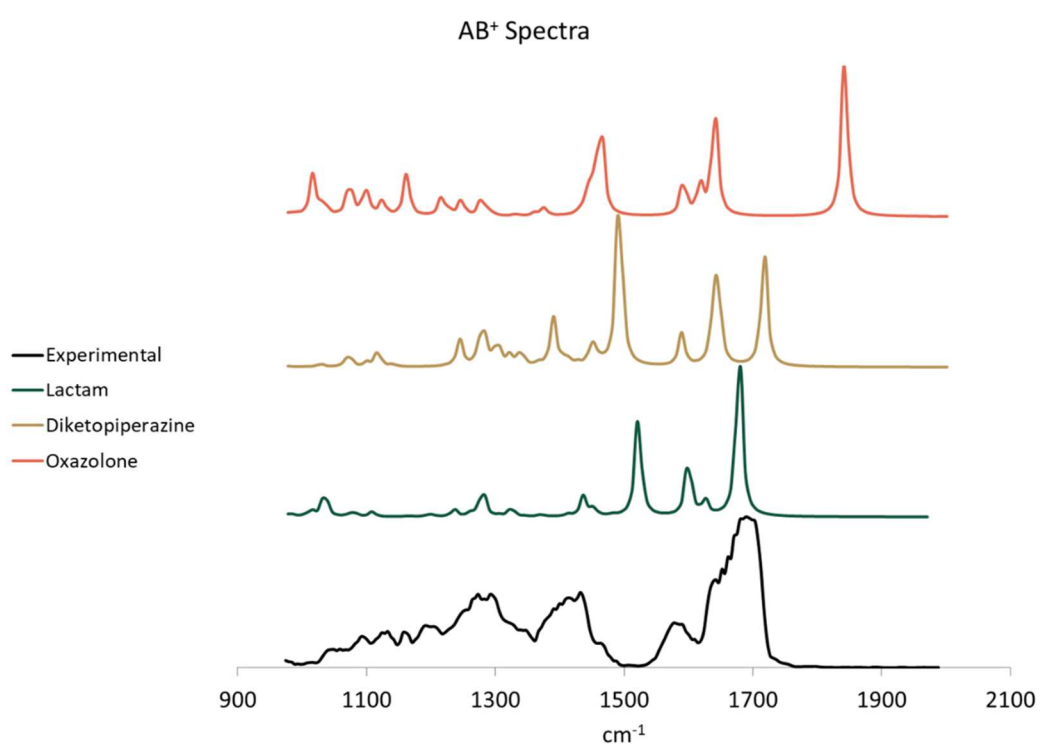


Figure 25 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AB^+.

The AB^+ experimental spectrum looks similar to that of AK^+ (See **Figure 26**). The broad distinct peak is blue-shifted by 50 cm^{-1} from the equivalent peak seen in the AK^+ spectrum. The AB^+ spectrum shows a low intensity but distinct peak near 1585 cm^{-1}, blue-shifted by ~10 cm^{-1} from the AK^+ spectrum. AB^+ shows the same loss of intensity near 1510 cm^{-1}. The lower energy peaks show similar features, though there is some distinct red shift in the AB^+ spectrum. The similarities indicate that AK^+ and AB^+ likely consist of the same isomer type or mixture of isomers.

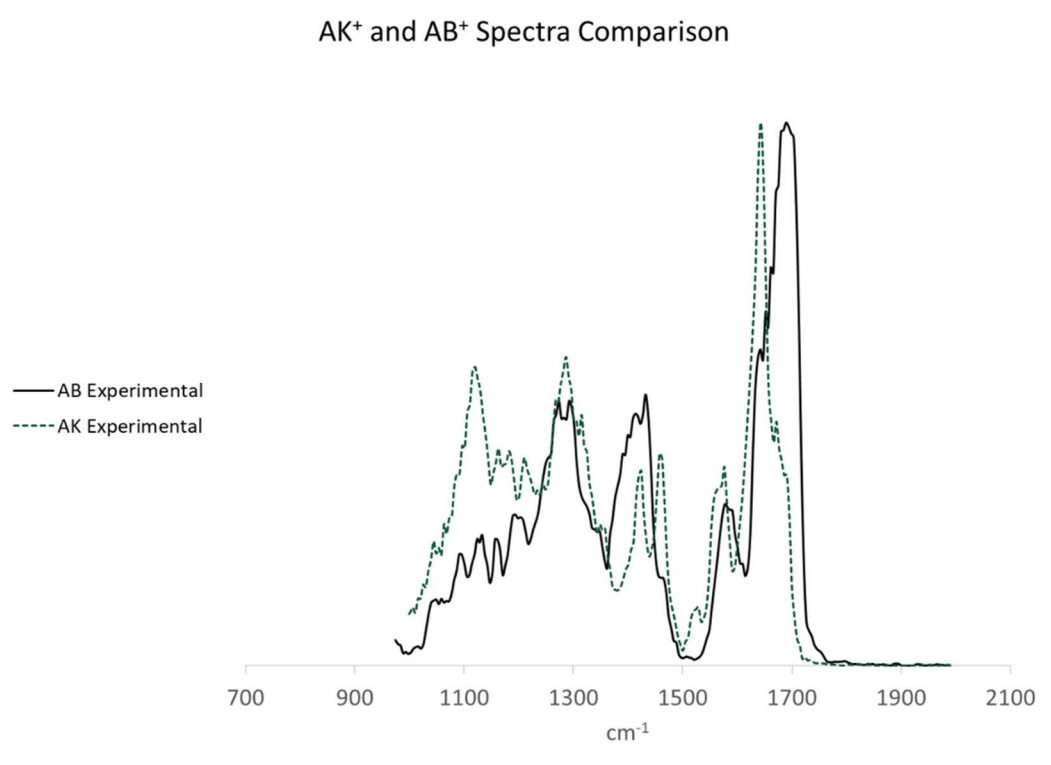


Figure 26 A spectral overlay of the experimental data captured at CLIO for both AB^+ (Solid Black) and AK^+ (Dashed Green).

Diaminopropionoic Acid Computational results for AZ^+ predict somewhat similar stabilities to the other b_2^+ fragments. Like AK^+, the lowest energy isomer was found to be the diketopiperazine protonated at the **Z** side chain of the **diket2** conformer type. The lactam isomer appears 56.4 kJ/mol higher in free energy than the diketopiperazine while the lowest lying oxazolone was calculated to be 65.2 kJ/mol higher in free energy. The lowest energy structure of each conformer type is tabulated here (See **Table 5**) and full computational results are reported in the Appendix (See **Table 19**).

Table 5 The lowest energy conformer of each isomer of AZ^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Conformer	ΔG^{298}	$\Delta\Delta G^{298}$
az_diket2_001	diket2	-550.3670408	0.0
az_diket1_018	diket1	-550.3582133	23.2
az_lac2_001	lac2	-550.3455517	56.4
az_diket3_013	diket3	-550.3448976	58.1
az_oxa1_017	oxa1	-550.3422206	65.2
az_lac3_001	lac2	-550.3328058	89.9
az_oxa3_009	oxa3	-550.3322373	91.4
az_oxa2_007	oxa2	-550.3267376	105.8
az_lac1_001	lac1	-550.3248632	110.7

The predicted lactam-diketopiperazine stability gap is much larger than that of AK^+ despite the similar isomer ordering. The sharp decrease in relative lactam stability as compared to the other b_2^+ fragments highlights how sterically disfavored the 4-membered lactam ring is. The relative increase in oxazolone stability (~15 kJ/mol) is likely due to the reduction of stability of the other two isomers rather than an inherent increase in the oxazolone stability. The shorter side chain length also reduces the available hydrogen bonding sites available to AZ^+. The two lowest diketopiperazine structures are of the **diket2** and **diket1** conformer types, respectively, and both show hydrogen bonding between the **Z** side chain and the **Z** carbonyl**.** Interestingly, when the proton shifts to the **Z** carbonyl, a conformer was not identified that presented that same hydrogen bonding scheme. Both lactam structures, understandably, present similar hydrogen bonding schemes to those seen in the other b_2^+ fragments as those schemes are not side chain length dependent. Hydrogen bonding schemes for other conformer types were maintained from the longer side chains (See **Figure 27**).

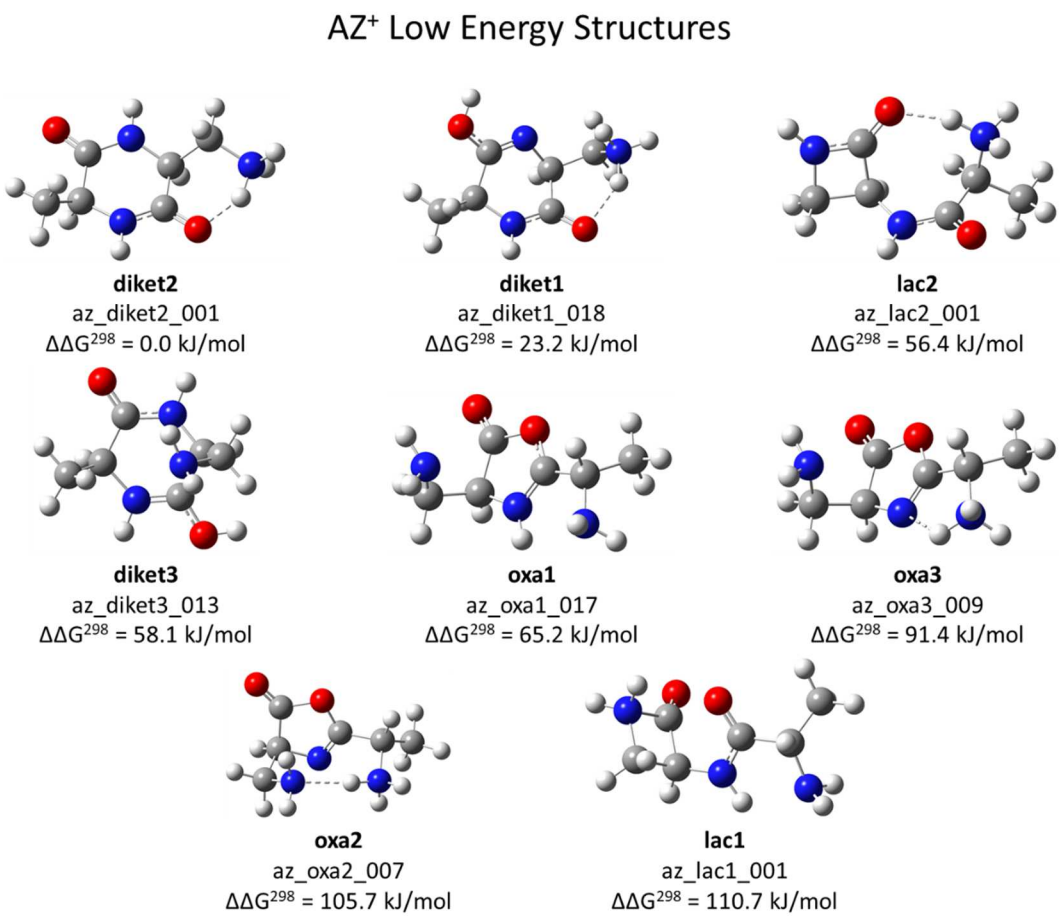


Figure 27 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AZ^+.

AZ^+ could not be isolated during the allotted laser time nor during subsequent hydrogen-deuterium exchange studies. As such, we do not have an experimental spectrum available for AZ^+. It is unclear whether AZAA forms a b_2^+ fragment. In the presence of lower energy fragmentation pathways, AZ^+ may be drowned out in the high intensity signal from other fragments or omitted entirely. While the existence of AAZ^+ confirms that **Z** terminal residues can form b-ions, however relatively unfavorable they may be, the presence of AZ^+ has yet to be confirmed in any of our mass spectrometer experiments.

Conclusion and Future Works

Infrared spectra in the fingerprint regions of AK^+, AO^+, AB^+, AK-lac^+, and AO-lac^+ were obtained using IRMPD action spectroscopy. The experimental spectra clearly indicate the generated fragments are not oxazolone isomers. Comparison of the AK^+ and AO^+ spectra to their novel lactam counterparts reveal differing results. The AK^+ fragment spectrum does not match the AK-lac^+ spectrum but does indicate some presence through mutually shared peaks. AO^+ and AO-lac^+ produced nearly identical spectra, confirming the AO^+ fragment is a lactam isomer. Comparison of the AB^+ and AK^+ spectra revealed similar absorption bands between the two, likely indicting shared structural features. As AK^+ is confirmed not to be an oxazolone and is not primarily a lactam, AK^+ is likely a mixture of the diketopiperazine and lactam forms. With no evidence for the formation of an AZ^+ b_2^+ fragment ion, future laser time should not be allotted to searching for a fragment that likely does not form.

Computational energetics corroborate the suggested results. AK^+ calculations show preferred stability for the diketopiperazine isomer (3.8 kJ/mol) over the lactam isomer. However, the relatively low stability gap indicates the two isomers could exist as a mixture from fragmentation. Interestingly, the AB^+ calculations show preferred stability for the lactam isomer (9.9 kJ/mol) against the diketopiperazine isomer, reversing the preference from AK^+. The closeness of the AK^+ and AB^+ spectra despite this stability shift suggests entropy might not be the driving factor in determining fragmentation pathway preference.

AO^+ clearly indicates that entropy does play a role in fragmentation pathway selection. A 6-membered lactam ring is less entropically expensive than a 6-membered diketopiperazine ring. This entropic cost reduction results in lactam isomers sufficiently stable enough that the diketopiperazine isomer is not observed in the experimental

spectrum. Likewise, a 5-membered lactam ring is more entropically favorable than the 6-membered diketopiperazine ring, as confirmed by the calculated stabilities. Yet the AB^+ experimental spectrum suggests some diketopiperazine population is present during fragmentation, despite the calculated energies. This suggests some other barrier plays a role in fragmentation pathway determination for the b_2^+ system.

Modelling of the fragmentation pathways is likely the immediate next step in describing this system. Identification of barriers to formation for the three isomers will better explain the observed fragment populations from the experimental spectrum. Barrier identification will likely involve the modelling of the intact tetrapeptide system as opposed to the products focused on in this study. A better understanding of the underlying principles governing the peptide's energetics will better instruct what product fragments are available. Additionally, novel diketopiperazine structures for AK^+ and AB^+ will be synthesized for their IRMPD action spectra. Neither AK^+ nor AB^+ predicted conclusive theoretical spectra to match the experimental data, and the use of novel structures was used with great success in identifying the structure of AO^+.

Chapter 5: b_3^+ Fragment Ions

Methods

Theoretical Procedures Calculations for the lactam, macrocycle, and oxazolone forms of AAK^+, AAO^+, AAB^+, and AAZ^+ were computed using the Gaussian 09 suite. Conformer sets were generated for several protonation sites (see **Figure 28**) using the GMMX searching routine in PCModel 9 which varies bond lengths and angles, saving structures with energies within 40 kJ/mol of the minimum energy structure. These conformer sets were imported into Gaussian and used as starting structures for a series of molecular orbital and density functional theory calculations. Final geometries, zero-point energies, and thermal corrections to enthalpy and free energies were computed and reported using the same procedure used for the b_2^+ fragment ions.

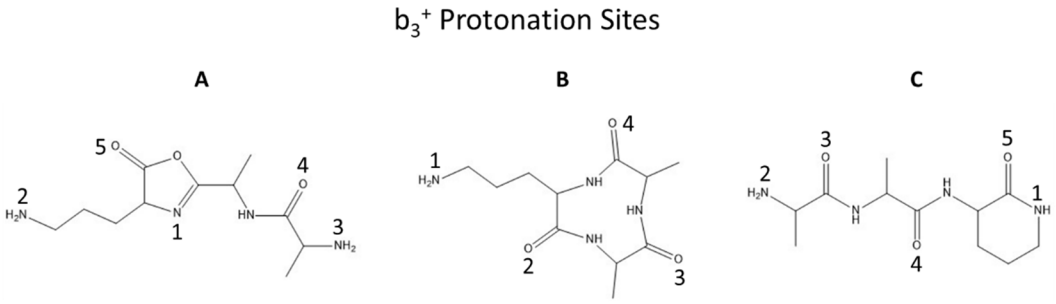


Figure 28 The b_3^+ fragment ions with ornithine in the terminal position are shown. **A** shows the oxazolone form, **B** shows the macrocyle form, and **C** shows the lactam form. The protonation sites' numbering does not imply an order of significance, rather just a naming scheme developed to differentiate conformers of the same isomer.

Naming of the b_3^+ fragment conformer types follows the convention set forth for the b_2^+ fragments (e.g. **lac4**). b_3^+ ions cannot form a diketopiperazine structure. They instead form a macrocycle structure, which will be shortened to **macro** in this text. The full word will be used when referring the isomer type while the shorthand will be used while discussing a specific conformer. The protonation site will use the naming scheme corresponding to the fragment type and might differ from the same number protonation

site for the other fragment. For example, a b_2^+ **Iac3** refers to a lactam protonated at the amide nitrogen while the same b_3^+ conformer refers to a lactam protonated at the first alanine carbonyl.

Experimental Procedures IRMPD spectra for the b_3^+ fragment ion species were obtained at using the FEL facility, CLIO, in Orsay, France and the FELIX facility in Nijmegen, the Netherlands. Infrared spectra for AAK^+, AOA^+, AAO^+, and AAB^+ were obtained in the fingerprint region using the CLIO FEL in a modified Bruker Esquire quadrupole ion trap mass spectrometer using the same conditions as described earlier. The AAZ^+ spectrum was obtained using the FELIX FEL in a modified Bruker AmaZon quadrupole ion trap mass spectrometer using the same conditions described earlier.

AAK^+ (*m/z* = 271 amu) fragments to yield peaks at *m/z* = 253 amu (loss of H_2O) and 128 which is likely the protonated remnant of the lysine residue (loss of $NH_3CHCH_3CONHCHCH_3CO$). AOA^+ (*m/z* = 257 amu) fragments to show peaks at *m/z* = 239 (loss of H_2O), 186 (loss of the terminal alanine residue), and 114 (loss of 143 amu shown from AAK^+). AAO^+ (*m/z* = 257 amu) observes the same fragment losses as AAK^+ (loss of 18 and 143 amu). AAB^+ (*m/z* = 243 amu) fragments to peaks at *m/z* 225 amu (loss of H_2O), 142 (loss of $NH_3CHCH_3CONHCH$), and 100 (loss of $NH_3CHCH_3CONHCHCH_3CO$). AAZ^+ (*m/z* = 229 amu) shows fragment peaks at *m/z* = 212 amu (loss of NH_3), 201 (loss of CO), 186 (loss of cyclic CH_2CH_2NH), and 158 (loss of $NH_2CHCH_3C{=}O$).

Materials Alanine Wang resin was purchased from ChemPep Inc (Wellington, FL). FMOC-Alanine-OH, FMOC-Lysine(BOC)-OH, FMOC-Ornithine(BOC)-OH, FMOC-diaminobutanoic acid(BOC)-OH, and FMOC-diaminopropionoic acid(BOC)-OH were purchased from MilliporeSigma (Burlington, MA). All materials were used without purification. Tetrapeptides were synthesized with lysine homologs in the third position

through the same solid-state synthesis procedure used in preparation of the b_2^+ tetrapeptides.

Results and Discussion

Comparison of the experimental spectra collected at CLIO and FELIX reveal general trends present among the amino side chains studied (See **Figure 29**). The AAK^+ and AAO^+ spectra follow very similar traces. They both present relatively sharp peaks of medium-low intensity around 1700 cm^{-1}, broad intense peaks near 1625 cm^{-1} with low energy shoulders, and then another discernable, broad absorption near 1440 cm^{-1}. In general terms, these spectra resemble that seen in the AK^+ experimental spectrum with the addition of the 1700 cm^{-1} peak, likely attributed to the additional carbonyl present in the third residue. The AAB^+ spectrum is defined by a single broad, high intensity absorption centered at 1680 cm^{-1}. A low intensity peak is shown near 1450 cm^{-1} and another of medium intensity near 1285 cm^{-1}. The AAZ^+ spectrum shows two peaks, one high intensity absorption at 1765 cm^{-1} with a high energy shoulder and another at 1680 cm^{-1}.

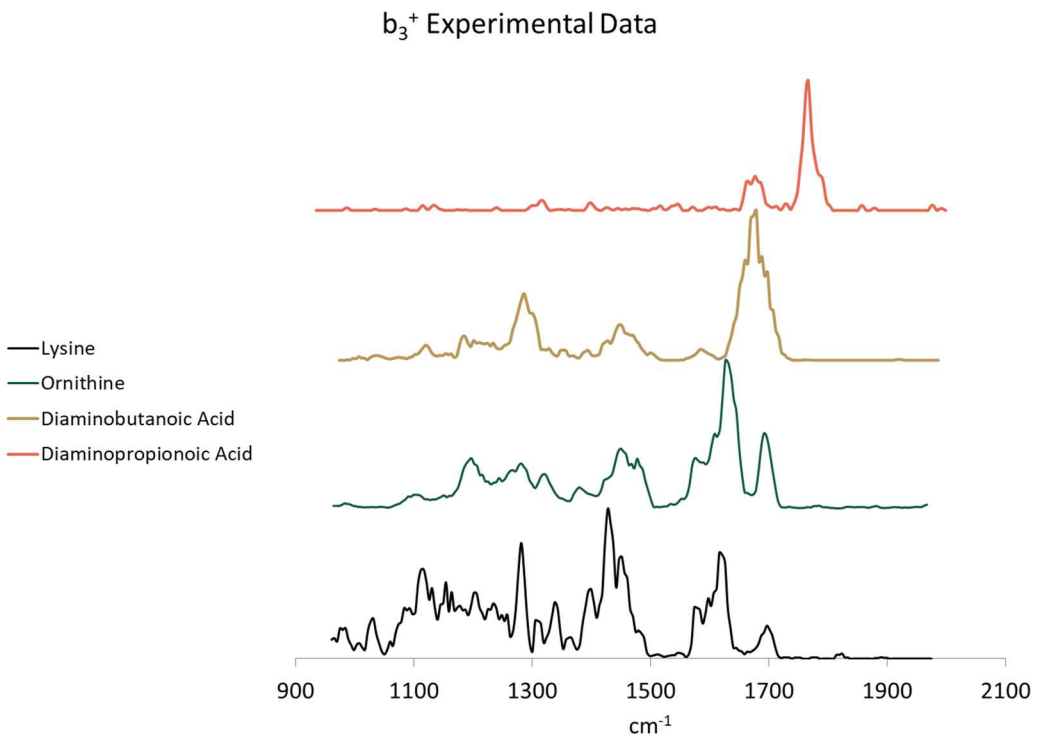


Figure 29 A comparison of the b_3^+ fragment IRMPD spectra collected at CLIO and FELIX shows AAK^+ (Black), AAO^+ (Green), AAB^+ (Gold), and AAZ^+ (Red).

This spectral analysis indicates AAK^+ and AAO^+ will have the same isomer type or a similar mixture of isomers. The change in spectral trace indicates AAB^+ consists of a different conformer type or isomer altogether. The AAZ^+ spectrum indicates a new conformer type as well. However, the peak at 1680 cm^{-1} suggests some of the AAB^+ conformer type may be present in the AAZ^+ spectrum as a mixture. Further analysis requires comparison with the computational results.

Lysine The lowest energy AAK^+ structure is a lactam isomer. The oxazolone and macrocycle are located 67.8 and 74.5 kJ/mol higher in free energy than the lactam global minimum. The relative stabilities of low lying conformers are tabulated here (See **Table 6**) and tabulated energy excerpts from the full computational search are available in the Appendix (See **Table 20**).

Table 6 The lowest energy conformer of each isomer of AAK^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Structure	ΔG^{298}	$\Delta\Delta G^{298}$
aak_lac3_123	lac2	-915.6127729	0.0
aak_lac5_154	lac3	-915.6123243	1.2
aak_lac5_202	lac5	-915.6068542	15.5
aak_lac5_113	lac4	-915.6009066	31.2
aak_lac3_211	lac6	-915.5912474	56.5
aak_lac4_034	lac7	-915.5869666	67.8
aak_oxa1_047	oxa1	-915.58694	67.8
aak_macro2_105	macro1	-915.5843853	74.5
aak_lac1_029	lac1	-915.5841357	75.2
aak_oxa5_060	oxa5	-915.5774322	92.8
aak_oxa4_024	oxa4	-915.5765556	95.1
aak_macro3_125	macro3	-915.5468572	173.1
aak_macro4_981	macro4	-915.5407005	189.2

The lowest energy AAK^+ lactam isomer is a **lac2**. The **lac2** conformer has a hydrogen bond between the protonated N-terminus and the ring carbonyl. This conformer type and hydrogen bonding scheme is maintained as the global minimum for each studied side chain, and therefore provides an excellent test of ring size on lactam stability. The move from b_2^+ to b_3^+ ions increased the 6-membered diketopiperazine structure to a 9-membered macrocyclic structure. The relative stability of this isomer type compared to the lactam dropped substantially due to the increased entropy cost of making a 9-membered ring.

Each of the conformer types for AAK^+ b_3^+ show unique hydrogen bonding schemes in their lowest energy geometries (See **Figure 30**). The lowest energy oxazolone, **oxa1**, developed a hydrogen bonding scheme where the protonated **K** side chain acts as a donor to both the ring nitrogen and the alanine carbonyl. The **macro1** conformer type adopted a hydrogen bonding scheme reminiscent of the **diket2** scheme—the protonated **K** side chain wraps back around to hydrogen bond with its nearest carbonyl.

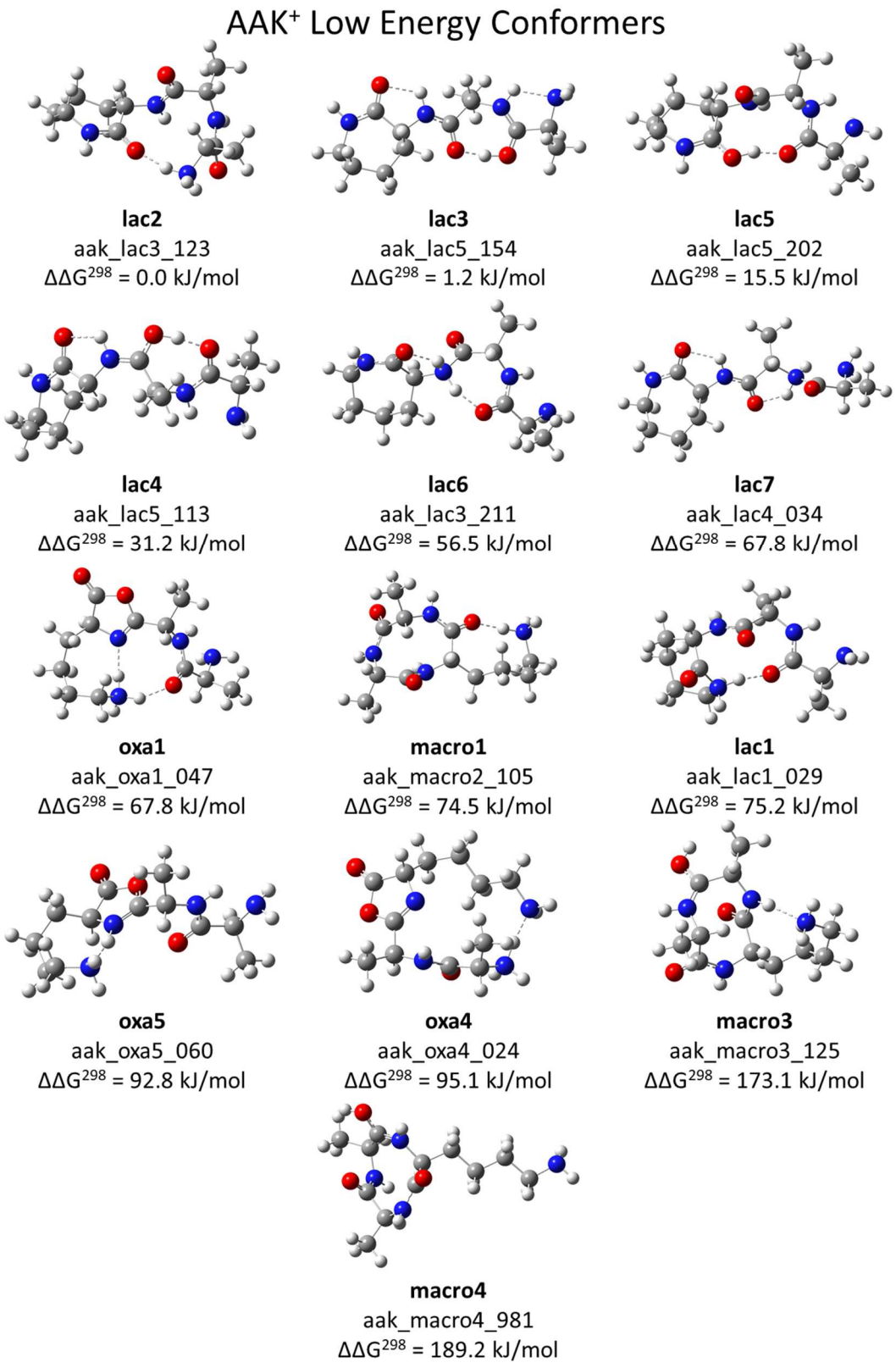


Figure 30 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AAK^+.

The spectra provided by these low energy conformers are compared to the experimental AAK^+ spectrum taken at CLIO (See **Figure 31**). The experimental spectrum matches well with the computed lactam spectrum. The most intense lactam peak, corresponding to the ring lactam, appears at 1625 cm^{-1} and matches the most intense band in the broad AAK^+ 1620 cm^{-1} absorption. The higher energy carbonyl peak also provides reasonable overlap with the lower intensity absorption in the experimental spectrum.

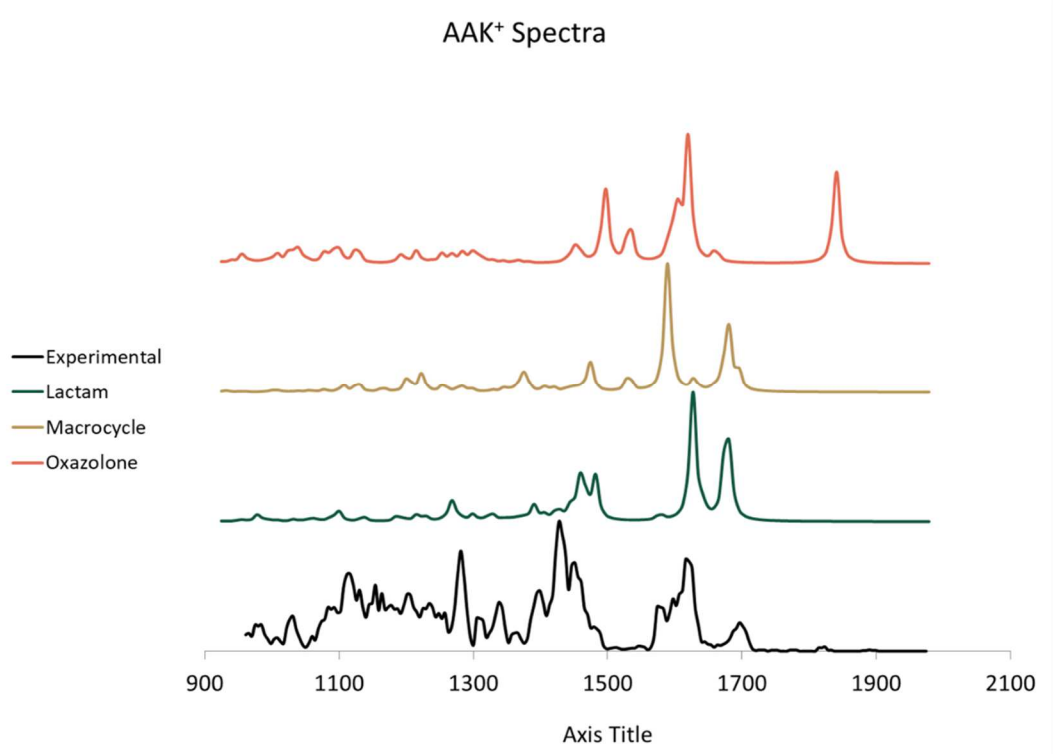


Figure 31 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AAK^+.

The broadening observed in the experimental spectrum may indicate some macrocycle presence. The low energy shoulder of the 1620 cm^{-1} experimental peak matches well with the intense macrocycle carbonyl stretch absorption at 1580 cm^{-1}. The macrocycle spectrum shares a broad peak with the lactam spectrum near 1680 cm^{-1}. This may account for the broadening seen in the experimental spectrum at that wavenumber.

As the AAK^+ and AAO^+ spectra share similar features, further insight will be drawn from the AAO^+ data.

Ornithine As shown in the b_2^+ computational search and IRMPD data, a 6-membered ring is the ideal size for the most stable lactam structure. As such, the lowest lying AAO^+ structure is a lactam. Like the other residues, the structure is the **lac2** conformer type. The oxazolone and macrocycle appear 73.6 and 86.5 kJ/mol higher in free energy. The relative stabilities of the various conformer types are tabulated here (See **Table 7**) and selections from the full computational search are available in the Appendix (See **Table 21**).

Table 7 The lowest energy conformer of each isomer of AAO^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Structure	ΔG^{298}	$\Delta\Delta G^{298}$
aao_lac2_001	lac2	-876.3186973	0.0
aao_lac5_027	lac4	-876.3175436	3.0
aao_lac3_084	lac3	-876.3145976	10.8
aao_lac5_012	lac5	-876.312686	15.8
aao_oxa3_628	oxa1	-876.2906606	73.6
aao_lac1_001	lac1	-876.2869252	83.4
aao_macro4_001	macro1	-876.2857432	86.5
aao_oxa5_074	oxa5	-876.2771303	109.1
aao_oxa4_078	oxa4	-876.2753332	113.9
aao_macro3_264	macro3	-876.2545174	168.5
aao_macro4_876	macro4	-876.2373395	213.6

As expected, the **lac2** conformer has increased relative stability compared to the similar AAK^+ structure provided by the smaller ring size. The **oxa1** conformer maintains the same hydrogen bonding scheme present in the AAK^+ **oxa1** and similar hydrogen bonding distances (ΔÅ < 0.02). Therefore, the decrease in the relative **oxa1** stability is due to the decrease in the lactam energy as opposed to a change in the oxazolone structure. The same does not hold true for the macrocyclic structures. The lowest energy conformer

is a **macro1** structure that shows a different hydrogen bonding scheme than the AAK^+ macrocycle structure. The AAO^+ **macro1** structure shows hydrogen bonding character between the protonated **O** side chain and the two nearest carbonyls. The AAK^+ macrocycle appeared 6.7 kJ/mol higher in free energy than the **oxa1** structure while the same gap for AAO^+ has increased to 12.9 kJ/mol. Despite the increased stability of an additional non-covalent interaction, the ring carbonyls have to orient themselves in a sterically unfavorable manner to produce these interactions. The result is a net stability decrease. The other conformers' hydrogen bonding schemes have been maintained from the AAK^+ structures (See **Figure 32**).

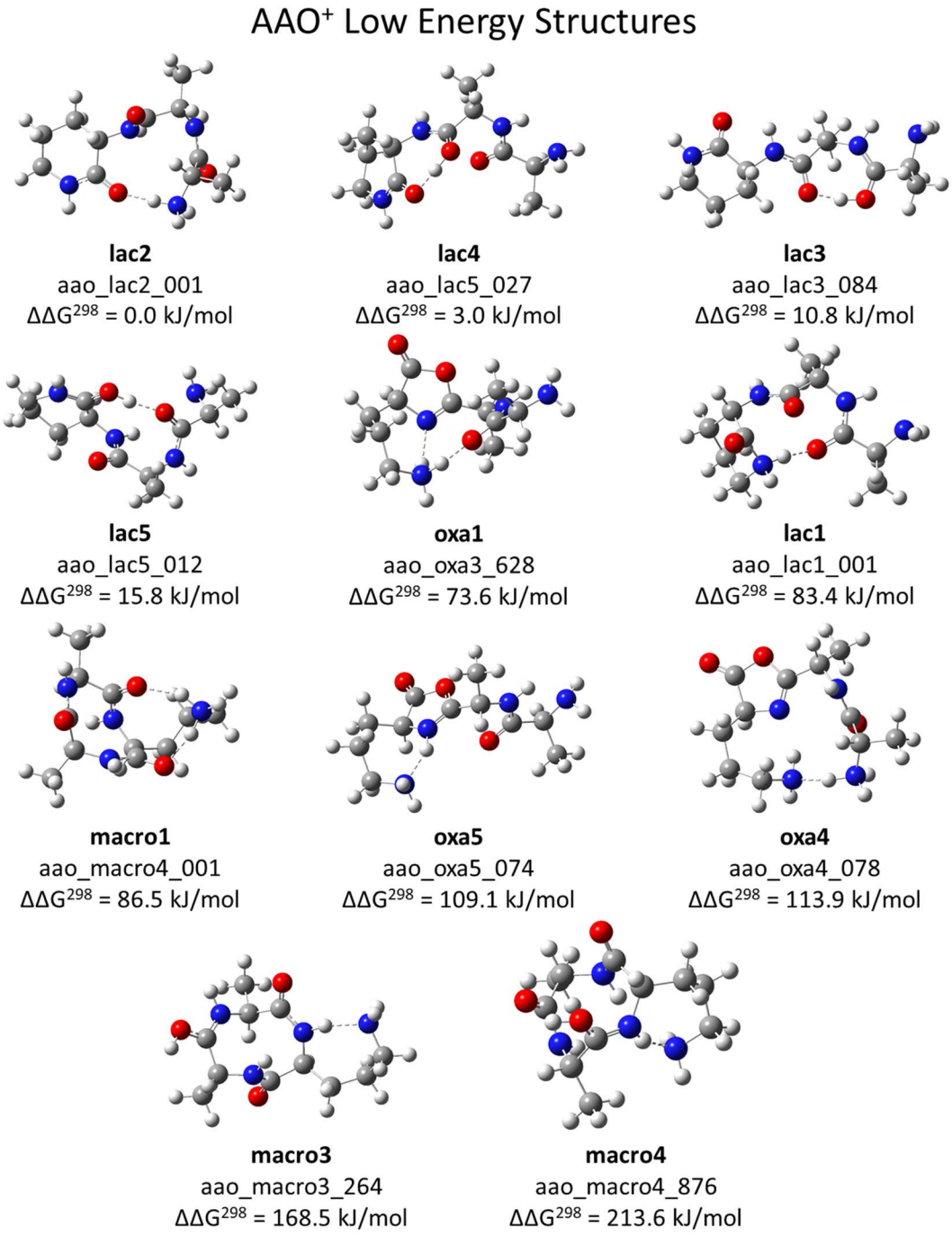


Figure 32 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AAO^+.

As indicated by the similarity of the AAK^+ and AAO^+ experimental spectra, the computed structures also produce similar predicted spectra (See **Figure 33**). The lactam isomer provides the best trace for the experimental spectrum as with AAK^+. The highest

intensity peaks of both spectra provide near perfect overlap cresting at 1627 cm^{-1} and 1630 cm^{-1} for the theoretical and experimental spectra, respectively. Additionally, the higher energy carbonyl stretch provides a match for the medium intensity absorbance at 1690 cm^{-1} in the experimental spectrum. The other pronounced peak 1482 cm^{-1} provides good overlap with the broad peak at 1480 cm^{-1} in the experimental spectrum.

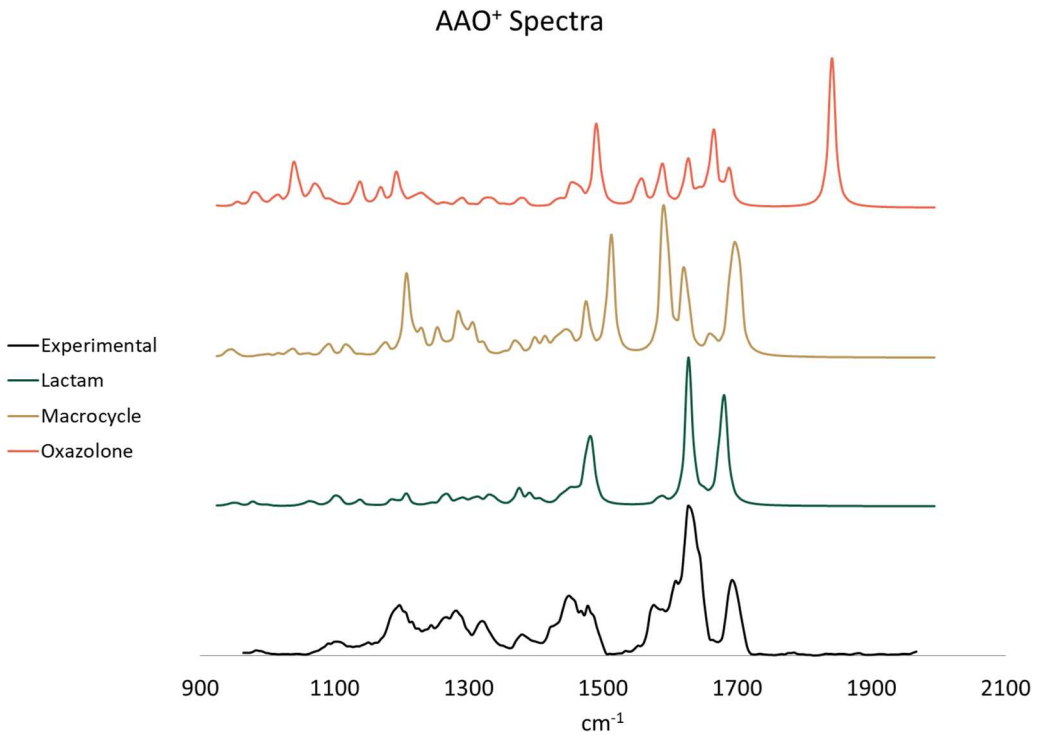


Figure 33 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AAO^+.

Unlike the AAK^+ spectrum, the theoretical macrocycle spectrum for AAO^+ differentiates itself from the experimental spectrum. A sharp, high intensity amide absorption at 1512 cm^{-1} in the theoretical spectrum falls into the region of the experimental spectrum lacking fragmentation. Further, the high intensity carbonyl stretch at 1595 cm^{-1} overlays an area of the spectrum with lower fragmentation than its surrounding signal, indicating no absorption save for the broadening effects.

An additional spectrum was taken of the AOA^+ fragment for comparison with the captured AAO^+ spectrum (See **Figure 34**). Because **O** is not the terminal residue for this fragment, the lactam channel is unavailable during fragmentation. Instead, the AOA^+ fragment can proceed through either the oxazolone or macrocycle pathways. Immediately noticeable is the lack of a strong oxazolone absorption in the AOA^+ spectrum. This fragment does not produce any distinguishable signal past 1750 cm^{-1}. This fragment must, therefore, be a macrocyclic compound. As the macrocycle is simply the cyclization of the N-terminus and the terminal residue's carbonyl carbon, the macrocycle for AAO^+ and AOA^+ should be identical—a 9-membered ring consisting of 3 carbonyls, 2 amide N-Hs, 2 alanine side chains, and 1 **O** side chain.

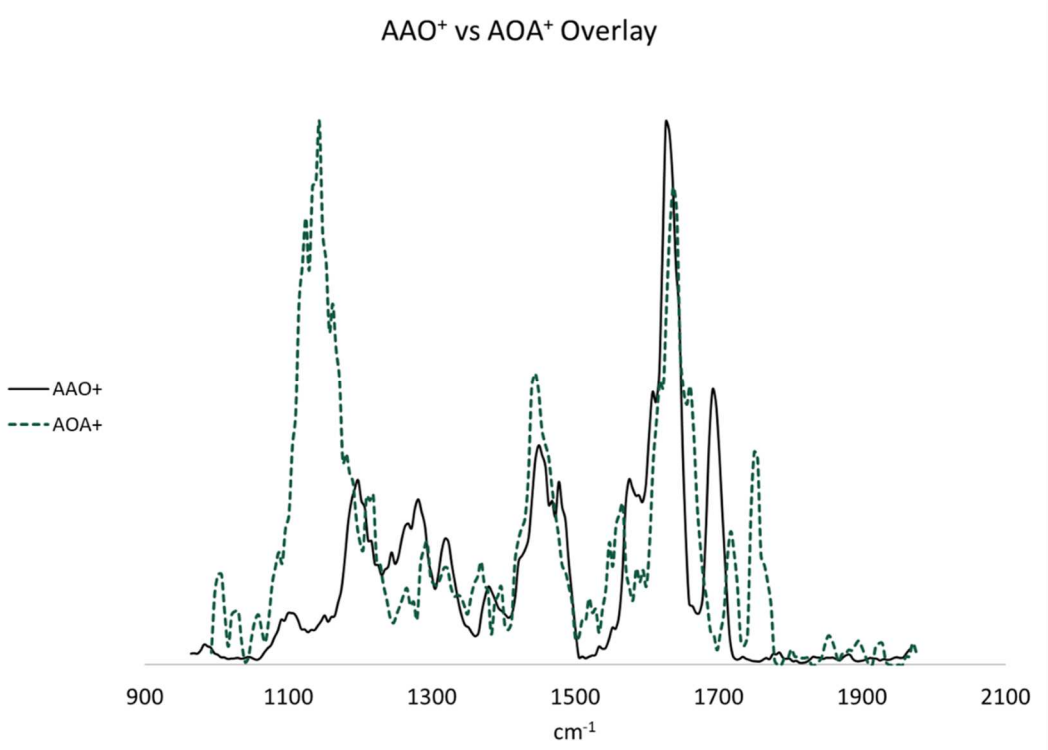


Figure 34 A spectral overlay of experimental data from AAO^+ (Solid Black) and AOA^+ (Dashed Green).

The overlay shows the similarities expected from an overlapping lactam and macrocycle spectrum. While the carbonyl stretches and NH_3 scissoring overlap well, two

higher energy peaks 1716 cm^{-1} and 1755 cm^{-1} as well as a low energy, high intensity 1143 cm^{-1} are completely absent in the AAO^+ spectrum. The clear differences confirm that AAO^+, and therefore AAK^+, is a lactam.

Diaminobutanoic Acid The lowest energy isomer of AAB^+ is a lactam of the **lac2** conformer type. The oxazolone and macrocycle isomers appear similarly high in free energy, at 85.4 and 86.2 kJ/mol, respectively. As shown in the other residues studied, the lowest lying conformer types are the **oxa1** and **macro1**. The relative stabilities of the lowest lying structure of each conformer type optimized is tabulated here (See **Table 8**) and selections from the full computational search are available in the Appendix (See **Table 22**).

Table 8 The lowest energy conformer of each isomer of AAB^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Structure	ΔG	ΔΔG
aab_lac1_0050	lac2	-837.0185456	0.0
aab_lac4_0101	lac4	-837.0140711	11.7
aab_lac3_0099	lac3	-837.0132205	14.0
aab_lac3_0039	lac5	-837.0113794	18.8
aab_oxa3_091	oxa1	-836.9860273	85.4
aab_macro2_249	macro1	-836.9857264	86.2
aab_lac1_0003	lac1	-836.9814889	97.3
aab_oxa5_006	oxa5	-836.9780881	106.2
aab_oxa2_025	oxa4	-836.9759523	111.8
aab_macro3_123	macro3	-836.9551557	166.4
aab_macro4_476	macro4	-836.9518235	175.2

The predicted AB^+ lactam isomer shows a similar hydrogen bonding scheme to those seen in the AO^+ and AK^+ lactam structures (See **Figure 35**). The **oxa1** conformer computed has a different hydrogen bonding scheme than previous **oxa1** structures. The AAB^+ conformer is not able to form a hydrogen bond with both the ring nitrogen and the

N-terminus, likely due to the reduced side chain length. Instead the **B** side chain shifts away from the ring nitrogen and shares the protonated amino group with the N-terminus. The loss of a hydrogen bond reduces the relative stability of the **oxa1** conformer quite dramatically (~11 kJ/mol). The reduction in stability places the **oxa1** conformer nearly iso-energetic with the **macro1** conformer. The **macro1** conformer adopts the hydrogen bonding scheme from AAK^+, but results in a similar relative stability to the AAO^+ scheme.

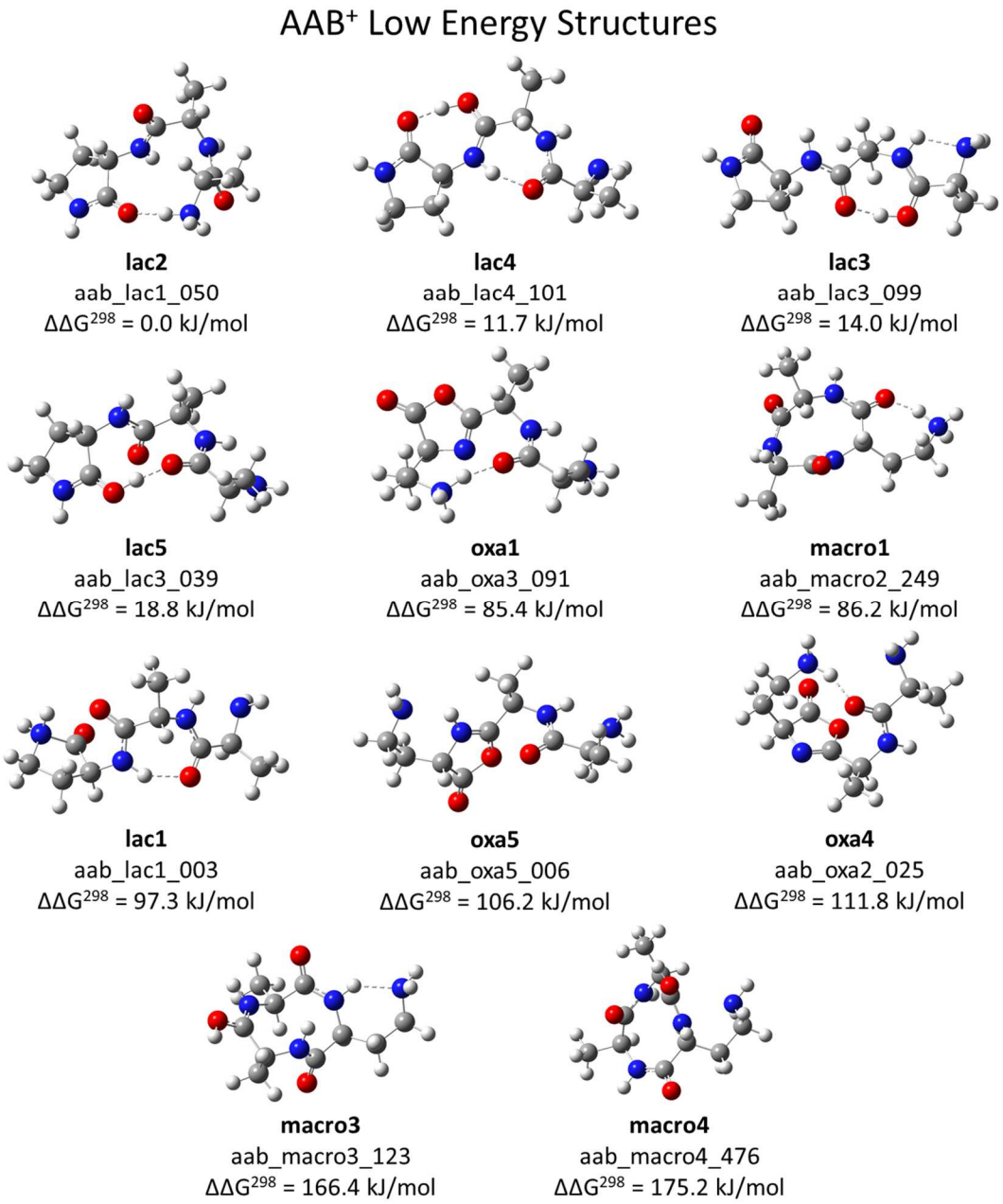


Figure 35 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AAB^+.

The theoretical spectra present different absorptions than those seen from the larger b_3^+ fragments (See **Figure 36**). Despite the similar hydrogen bonding scheme present in the **lac2** structure, the spectrum varies from that seen in the two larger isomers. There is near complete overlap of the three carbonyl stretching modes in a relatively broad, intense band near 1680 cm^{-1}. Similarly, there is significant overlap of the amide

stretching modes and NH_3 rocking mode that combine at a single relatively sharp, medium intensity band at 1480 cm^{-1}. A series of low intensity absorptions litter the low energy portion of the spectrum, the most notable of which is a set of overlapping, low intensity CH_2 wagging and twisting modes near 1270 cm^{-1}.

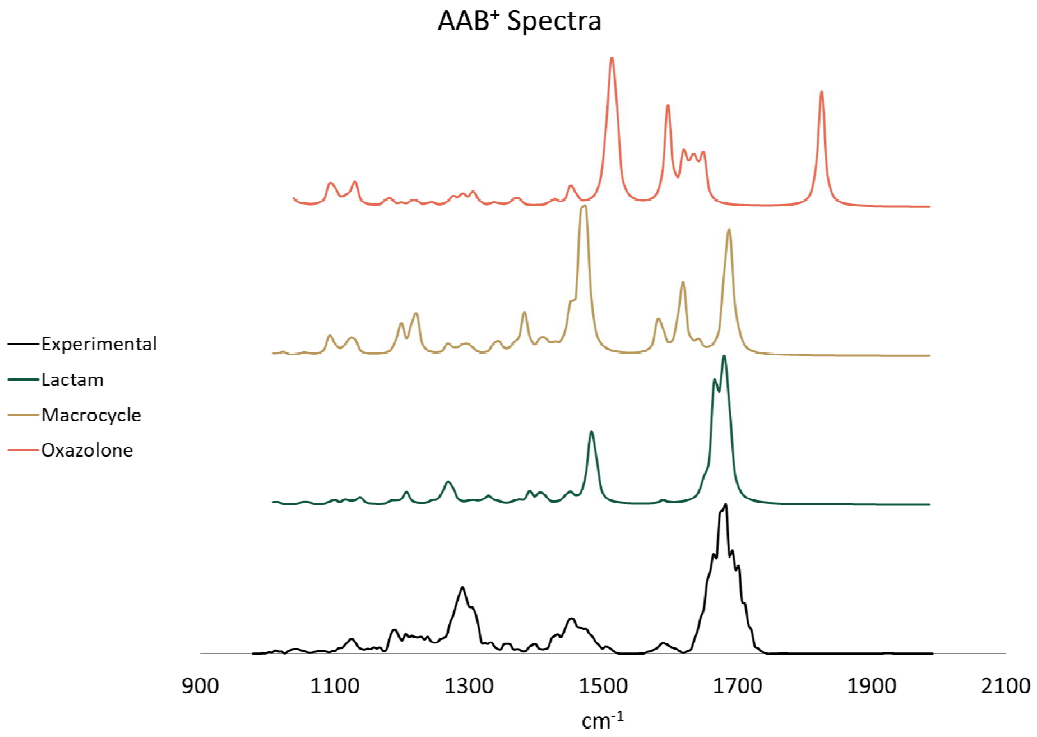


Figure 36 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AAB^+.

As noted in the experimental spectra comparison, the AAB^+ fragment spectrum differed from that of AAK^+ and AAO^+. However, the lactam spectrum clearly provides the best trace for the experimental data observed. The large, broad absorption at 1680 cm^{-1} matches the overlapping carbonyl stretches from the **lac2** spectrum. The experimental spectrum predicts a very low absorption near 1595 cm^{-1} that is faintly visible in the **lac2** spectrum, corresponding to the N-terminus bending mode. The 1480 cm^{-1} absorption in the **lac2** spectrum appears red-shifted by ~20 cm^{-1} in the experimental spectrum. However, there are discernable peaks at the corresponding wavenumber in either

spectrum, potentially indicating that the lower absorbing CH_2 bending modes from the **Iac2** spectrum provided better infrared absorbance and fragmentation in the experimental spectrum.

In addition to the similarities shown between the **Iac2** and experimental spectra, the macrocycle and oxazolone spectra both show distinct differences that support a single isomer fragment theory. The oxazolone, as shown in each other spectrum, predicts a sharp, high energy absorption 1826 cm^{-1} corresponding to the ring carbonyl. This peak is clearly not present in the experimental spectrum. The macrocycle shows spectral features more generally agreeable with the experimental spectrum than the oxazolone spectrum did. However, the sharp, high intensity carbonyl absorption at 1688 cm^{-1} appears slightly red shifted than the broad absorption in the experimental spectrum, and, more importantly, lacks an absorption immediately lower in energy that might imply the broadening seen in the experimental spectrum. Further, while the macrocycle spectrum predicts two absorptions (1619 cm^{-1} and 1581 cm^{-1}) near the experimental peak at 1596 cm^{-1}, it again lacks any signal between those peaks to suggest the absorption seen in the experimental spectrum.

Diaminopropionoic Acid Like the other b_3^+ fragments studied, the lowest energy AAZ^+ isomer was computed to be a lactam protonated at the N-terminus. As shown in the AZ^+ calculated structures, the 4-member lactam ring is not as energetically favorable as its higher order counterparts. As such, AAZ^+ displays condensed stability gaps as compared to the other b_3^+ fragments. The macrocycle appears 15.4 kJ/mol higher while the oxazolone has a $\Delta\Delta G$ of 19.2 kJ/mol. The lowest energy structure of each conformer type from the computational search are reported below (See **Table 9**). Excerpts from the full computational results are available in the Appendix (See **Table 23**).

Table 9 The lowest energy conformer of each isomer of AAZ^+ is shown. ΔG values are reported in Hartrees while ΔΔG values have been converted to kJ/mol.

Name	Structure	ΔG	ΔΔG
aaz_lac2_0002	lac2	-797.6865439	0.0
aaz_macro4_0305	macro1	-797.6806928	15.4
aaz_lac3_0040	lac3	-797.6803864	16.2
aaz_oxa1_0021	oxa1	-797.6792261	19.2
aaz_lac5_0010	lac4	-797.6767867	25.6
aaz_oxa5_0004	oxa5	-797.6759512	27.8
aaz_lac3_0116	lac5	-797.6725442	36.8
aaz_oxa1_0015	oxa4	-797.671787	38.7
aaz_macro3_0075	macro3	-797.654594	83.9
aaz_macro3_0089	macro4	-797.6406571	120.5
aaz_macro3_0021	macro2	-797.6367833	130.6

As shown in the other b_3^+ structures, the lowest energy structure is of the **lac2** conformer type. This structure shares the same hydrogen bonding scheme expected from a **lac2** conformer. The macrocyclic structure is of the **macro1** conformer type. Interestingly, this **macro1** adopts the hydrogen bonding scheme shown in the AAO^+ structure as opposed those seen in AAK^+ and AAB^+. The steric hindrance of the lactam structure increases the stability of the macrocycle structure by ~60 kJ/mol relative to the AAO^+ stabilities. Unique to the AAZ^+, the macrocycle appears more stable than the oxazolone. This change is likely due to the shortened **Z** side chain increasing the hydrogen bonding distance in the **oxa1** conformer. The increase in hydrogen bonding length reduces the stabilizing effect of that non-covalent interaction. Further, the AAZ^+ **macro1** has additional hydrogen bonding character not present in the AAB^+ **macro1** which already had a similar stability to the **oxa1** conformer (+0.8 kJ/mol). The removal of a methylene from the amino side chain was enough to flip the stabilities of those two isomers.

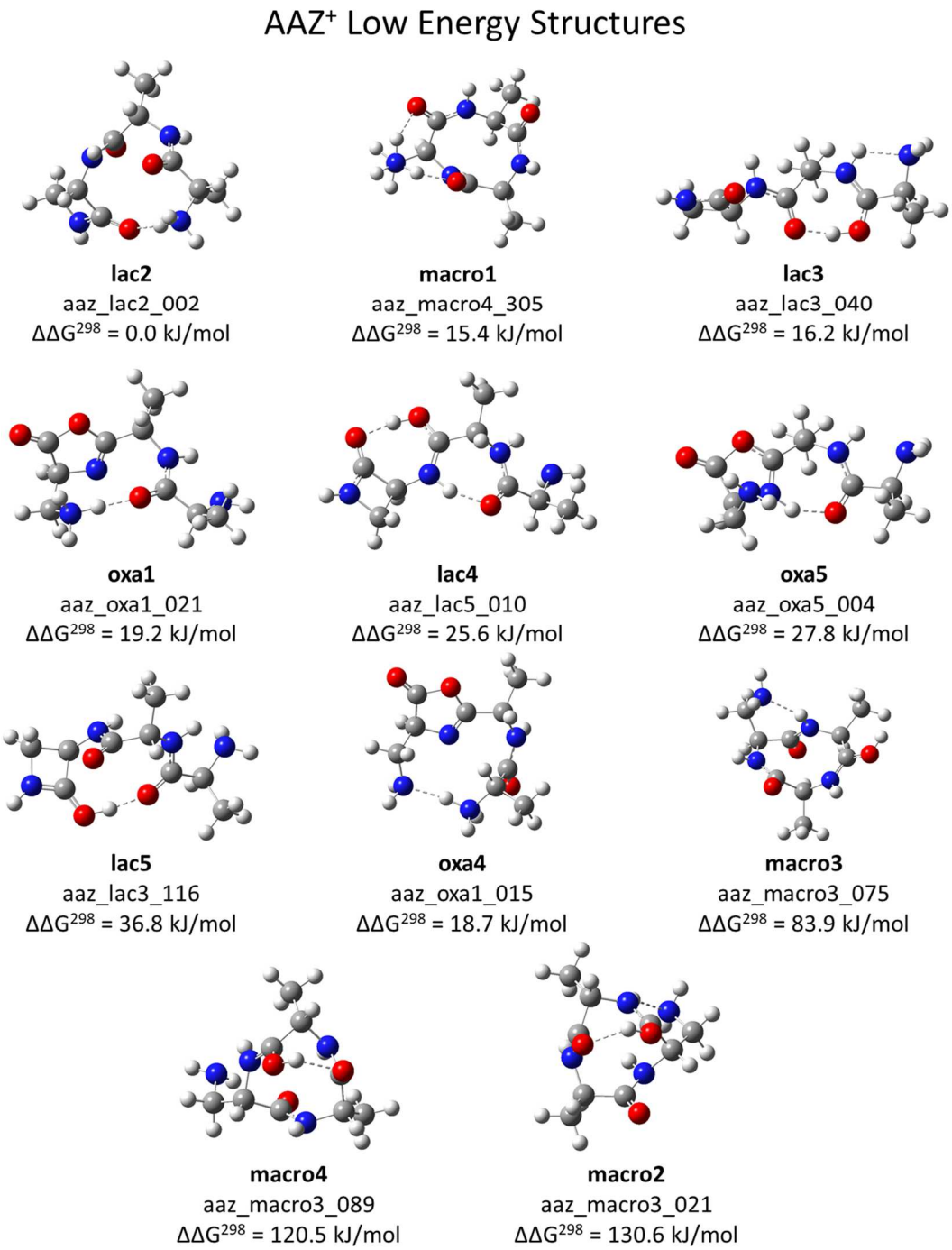


Figure 37 B3LYP/6-31+G(d) optimized structures for the lowest energy conformer type of each isomer of AAZ^+.

AAZ^+ was the only studied residue to optimize a **macro2** conformer (See **Figure 37**). Though it appears significantly higher in energy (130.6 kJ/mol), this structure presents a new hydrogen bonding scheme characterized by a 5-membered hydrogen bonding ring

between the **Z** side chain and the amide hydrogen as well as the protonated carbonyl sharing its proton with the adjacent carbonyl. For further determination of stability dependent geometries, we turn to the spectra.

The experimental spectrum appears wholly distinct from the predicted theoretical spectra (See **Figure 38**). There were two problems in acquiring data that likely resulted in a poor experimental spectrum. First, the b_3^+ fragment from AAZA was not the preferred pathway. CID activation showed the most intense peaks at *m/z* = 300 and 176, corresponding to the loss of water and the y_2^+ fragment. The b_3^+ was isolated, but two orders of magnitude of signal were lost during fragmentation. A reduced population will give less accurate results, and the sharp drop between MS/MS and MS^3 results in the trap trying to compensate for the lack of ions.

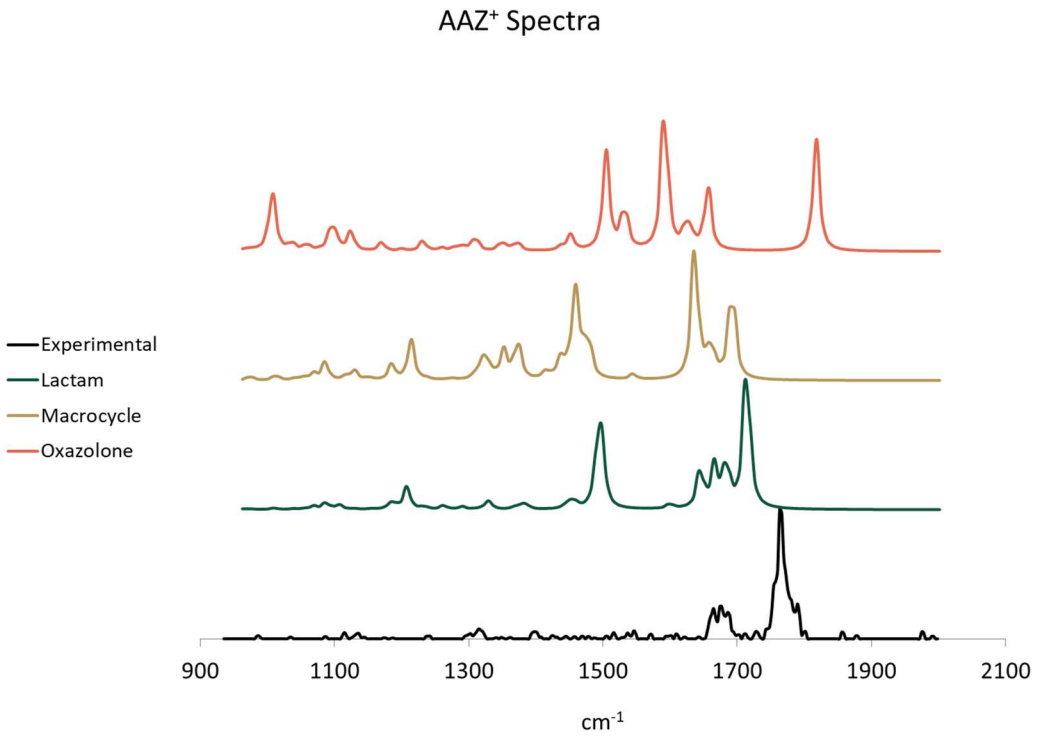


Figure 38 A comparison of the experimental IRMPD spectrum (Black) and the theoretical spectra for the lowest energy lactam (Green), diketopiperazine (Gold), and oxazolone (Red) forms of AAB^+.

Further, the laser was operating on significantly less power during this time. The power in the higher energy region of the spectrum was reduced to a sixth of the normal output, though slight increases were gained as the wavelength increased. Reduced power results in fewer total photons entering the trap, and, as the IRMPD pathway relies on the absorption of multiple photons, less fragmentation. Coupled with the reduced signal, this AAZ^+ is likely only accurate as a measure of the most intense absorbers. We attempted to mitigate the reduction in power by increasing the number of micropulses and averages but did not see a discernable increase in fragmentation efficiency. The reported spectrum is the result of 4 micropulses and 6 averages. The resulting flat spectrum likely has much more detail available to a more powerful laser but was washed out by the lack of fragmentation.

Comparison to this spectrum will not be accurate per se, though the two peaks observed in the experimental spectrum are certainly there. The theoretical lactam spectrum shows a similarly barren absorption range. The broad triplet-like peak centered around 1680 cm^{-1} provides good overlap with the same trace seen in the experimental spectrum. The lactam carbonyl stretch at 1710 cm^{-1} shows the same shape and intensity as the AAZ^+ carbonyl, but appears 50 cm^{-1} too low in energy. The amide stretch near 1500 cm^{-1} is not shown in the experimental spectrum whatsoever. The diketopiperazine and oxazolone both predict similar carbonyl traces near 1700 cm^{-1} and 1685 cm^{-1}, however, these both appear significantly red-shifted as compared to the experimental spectrum. The AAZ^+ carbonyl stretch is not quite blue-shifted enough to look like the oxazolone structure. If a lower intensity oxazolone carbonyl or amide stretch were present in the AAZ^+ spectrum, it was likely entirely washed out. As it stands none of the theoretical spectra match that of the experimental. However, another scan with increased laser power might coax out more detail.

Conclusion and Future Work

Infrared spectra in the fingerprint region of b_3^+ fragment ions with **K**, **O**, **B**, and **Z** in the terminal position were obtained using IRMPD action spectroscopy. In contrast to the b_2^+ ions, which showed inconclusive structural determination, we present clear spectroscopic evidence that AAK^+, AAO^+, and AAB^+ form a lactam b_3^+ fragment. These experimental spectra are supported by robust computational methods that also predict the AAZ^+ fragment will form a lactam b_3^+, though experimental data gathered for this species remains inconclusive.

The shift in favor of lactam stability as compared to the b_2^+ system is likely a result of the increased ring size of the diketopiperazine corollary—the 9-membered macrocycle ring. While the b_2^+ system estimated the isomeric stability of the lactam as compared to the diketopiperazine to be < ~15 kJ/mol for all species while the shift to a macrocycle increased that stability to > 70 kJ/mol in the b_3^+ system. The exceptions for both systems are of course the **Z** containing fragments. The entropic cost of creating the 4-membered lactam ring significantly reduces the stability of that isomer, though it is still favored (15.4 kJ/mol) in the b_3^+ system.

Therefore, entropy does seem to play a significant role in the fragmentation pathway for the b_3^+ ions. The entropic cost of organizing a ring of n atoms will decrease with n (until n = 4). In addition to its large size (n = 9), the macrocycle is also fighting against partial double bond character in the amide bonds in fragment formation, further reducing its stability. This results in a fragment overwhelmingly favoring the lactam pathway during its cleavage.

The AAZ^+ spectrum will need to be retaken at a time when the FEL is operating at full power output. The gathered spectrum appeared suppressed, and further detail will

likely be revealed if additional fragment intensity is observed. Short of increased power, perhaps longer periods of signal averaging would also help. Allotted laser time will always be an issue in allocating infrared resources, but the captured spectrum provides evidence that the fragment is formed and will require more work for a full picture.

Appendix

Tables

Table 10 Full computational results for the M06-2X/6-311++G(d,p) computational search of AK^+. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ak_diket1_077_m06	diket2	-667.988109	0.0
ak_lac4_024_trans_m06	lac5	-667.986487	4.3
ak_lac2_003_m06	lac2	-667.981917	16.3
ak_lac4_024_m06	lac5	-667.973158	39.3
ak_lac3_002_m06	lac3	-667.962743	66.6
ak_lac1_001_m06	lac1	-667.961765	69.2
ak_lac3_122_m06	lac4	-667.96121	70.6
ak_oxa3_017_oxa2_m06	oxa2	-667.959219	75.8
ak_oxa3_017_m06	oxa2	-667.956155	83.9
ak_diket2_107_m06	diket3	-667.954125	89.2
ak_oxa3_069_m06	oxa3	-667.954019	89.5
ak_ox_m06_056_m06	oxa1	-667.944818	113.7
ak_diket3_021_m06	diket1	-667.938971	129.0

Table 11 Full computational results for the ωB97-XD/6-311++G(d,p) computational search of AK^+. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ak_diket1_077_wb	diket2	-668.074917	0.0
ak_lac4_024_trans_wb	lac5	-668.071363	9.3
ak_lac2_003_wb	lac2	-668.070096	12.7
ak_lac4_024_wb	lac5	-668.058173	44.0
ak_lac3_002_wb	lac3	-668.049096	67.8
ak_lac1_001_wb	lac1	-668.047656	71.6
ak_oxa3_017_oxa2_wb	oxa2	-668.044754	79.2
ak_diket2_107_wb	diket3	-668.041213	88.5
ak_oxa3_069_wb	oxa3	-668.039445	93.1
ak_ox_wb_056_wb	oxa1	-668.032423	111.6
ak_oxa3_017_wb	oxa3	-668.031416	114.2
ak_lac3_122_wb	lac3	-668.027122	125.5
ak_diket3_021_wb	diket1	-668.025465	129.8

Table 12 Full computational results for the M06-2X/6-311++G(d,p) computational search of AO^+. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ao_lac4_012_trans_m06	lac5	-628.712973	0.0
ao_lac2_001_m06	lac2	-628.710607	6.2
ao_diket2_010_m06	diket2	-628.705952	18.4
ao_lac4_012_m06	lac5	-628.701985	28.8
ao_diket1_305_m06	diket2	-628.690684	58.5
ao_diket1_305_diket4_m06	diket4	-628.690684	58.5
ao_lac1_005_m06	lac5	-628.690171	59.9
ao_lac3_002_m06	lac3	-628.682511	80.0
ao_oxa2_001_m06	oxa2	-628.68134	83.1
ao_diket3_360_m06	diket3	-628.676807	95.0
ao_oxa3_069_m06	oxa3	-628.670195	112.3
ao_oxa1_052_m06	oxa2	-628.669708	113.6
ao_diket1_286_m06	diket1	-628.669216	114.9
ao_oxa1_165_m06	oxa1	-628.667453	119.5
ao_diket1_139_m06	diket1	-628.658861	142.1
ao_diket3_339_m06	diket3	-628.657511	145.6

Table 13 Full computational results for the ωB97-XD/6-311++G(d,p) computational search of AO^+. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ao_lac4_012_trans_wb	lac5	-628.789271	0.0
ao_lac2_001_wb	lac2	-628.787951	3.5
ao_diket2_010_wb	diket2	-628.783476	15.2
ao_lac4_012_wb	lac5	-628.775473	36.2
ao_diket1_305_diket4_wb	diket4	-628.766822	58.9
ao_diket1_305_wb	diket1	-628.765725	61.8
ao_lac1_005_wb	lac1	-628.762422	70.5
ao_lac3_002_wb	lac3	-628.760475	75.6
ao_oxa2_001_wb	oxa2	-628.757373	83.7
ao_diket3_360_wb	diket3	-628.75256	96.4
ao_ox_wb_052_wb	oxa1	-628.74566	114.5
ao_diket1_286_wb	diket1	-628.745208	115.7
ao_oxa3_069_wb	oxa3	-628.744528	117.5
ao_ox_wb_165_wb	oxa1	-628.74148	125.5
ao_diket3_339_wb	diket3	-628.737223	136.6
ao_diket1_139_wb	diket1	-628.735621	140.9

Table 14 Full computational results for the M06-2X/6-311++G(d,p) computational search of AB^+. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ab_lac1_001_lac5_m06	lac5	-589.429753	0.0
ab_diket2_001_m06	diket2	-589.425691	10.7
ab_diket3_009_diket2_m06	diket2	-589.425678	10.7
ab_lac2_007_m06	lac2	-589.417118	33.2
ab_diket1_071_diket4_m06	diket4	-589.408195	56.6
ab_diket1_071_m06	diket1	-589.405722	63.1
ab_diket3_130_m06	diket3	-589.401617	73.9
ab_lac1_001_m06	lac1	-589.3986	81.8
ab_lac3_007_m06	lac3	-589.396173	88.2
ab_oxa1_004_m06	oxa1	-589.393537	95.1
ab_diket3_009_m06	diket3	-589.393448	95.3
ab_oxa3_003_oxa2_m06	oxa2	-589.393206	96.0
ab_diket1_084_m06	diket1	-589.387775	110.2
ab_oxa1_030_m06	oxa1	-589.384185	119.6
ab_oxa3_003_m06	oxa3	-589.382669	123.6
ab_oxa3_005_m06	oxa3	-589.377487	137.2

Table 15 Full computational results for the ωB97-XD/6-311++G(d,p) computational search of AB^+. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ab_lac2_007_wb	lac2	-589.498242	0.0
ab_lac1_001_lac5_wb	lac5	-589.497448	2.1
ab_diket3_009_wb	diket2	-589.495394	7.5
ab_diket2_001_wb	diket2	-589.495392	7.5
ab_diket1_071_diket4_m06	diket4	-589.474564	62.2
ab_diket1_071_wb	diket1	-589.47327	65.6
ab_lac1_001_wb	lac1	-589.46675	82.7
ab_diket3_130_wb	diket3	-589.466393	83.6
ab_lac3_007_wb	lac3	-589.464935	87.4
ab_oxa3_003_oxa2_wb	oxa2	-589.4594	102.0
ab_oxa1_004_wb	oxa1	-589.459327	102.2
ab_oxa1_004_oxa1_wb	oxa1	-589.45932	102.2
ab_oxa1_030_wb	oxa1	-589.456657	109.2
ab_diket1_084_wb	diket1	-589.454799	114.1
ab_oxa3_003_wb	oxa3	-589.449262	128.6
ab_oxa3_005_wb	oxa3	-589.443311	144.2

Table 16 The 60 lowest energy AK^+ conformers out of 440 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ak_diket_077	diket2	-668.2778117	0.0
ak_diket_006	diket2	-668.277371	1.2
ak_diket_083	diket2	-668.2771276	1.8
ak_diket_071	diket2	-668.276552	3.3
ak_lac4_024	lac4	-668.2763483	3.8
ak_diket_040	diket2	-668.275494	6.1
ak_diket_009	diket2	-668.2754593	6.2
ak_diket_018	diket2	-668.275033	7.3
ak_diket_038	diket2	-668.2743763	9.0
ak_lac4_043	lac2	-668.274315	9.2
ak_lac2_003	lac2	-668.2743106	9.2
ak_diket_043	diket2	-668.2739255	10.2
ak_diket_003	diket2	-668.2728426	13.0
ak_diket_027	diket2	-668.2727679	13.2
ak_lac4_013	lac4	-668.2718071	15.8
ak_lac4_041	lac4	-668.2712263	17.3
ak_lac2_001	lac2	-668.2708848	18.2
ak_lac4_019	lac4	-668.270254	19.8
ak_lac4_044	lac4	-668.2700113	20.5
ak_lac4_018	lac2	-668.2698789	20.8
ak_lac2_002	lac2	-668.2698129	21.0
ak_lac4_008	lac2	-668.2698116	21.0
ak_lac4_022	lac4	-668.2693358	22.3
ak_lac4_034	lac2	-668.2689874	23.2
ak_lac2_012	lac2	-668.2689791	23.2
ak_lac4_051	lac4	-668.2677185	26.5
ak_lac2_010	lac2	-668.2670405	28.3
ak_lac_018	lac2	-668.2659708	31.1
ak_lac3_153	lac2	-668.2659671	31.1
ak_lac2_017	lac2	-668.2656688	31.9
ak_lac2_009	lac2	-668.2647465	34.3
ak_lac2_007	lac2	-668.2646149	34.6
ak_lac4_029	lac2	-668.2644424	35.1
ak_lac2_013	lac2	-668.2644356	35.1
ak_lac4_025	lac2	-668.2641369	35.9
ak_lac2_004	lac2	-668.2641274	35.9
ak_lac4_001	lac4	-668.2638848	36.6
ak_lac4_048	lac2	-668.2605465	45.3

ak_lac4_047	lac4	-668.2596693	47.6
ak_lac4_010	lac5	-668.2590484	49.3
ak_lac4_009	lac5	-668.2589714	49.5
ak_diket_032	diket2	-668.2582444	51.4
ak_lac2_014	lac2	-668.2579588	52.1
ak_lac4_027	lac2	-668.2579086	52.3
ak_lac4_007	lac5	-668.2578933	52.3
ak_lac4_014	lac5	-668.2575599	53.2
ak_diket_060	diket2	-668.2575137	53.3
ak_lac2_016	lac2	-668.2568135	55.1
ak_diket_089	diket2	-668.2565103	55.9
ak_diket_094	diket2	-668.256413	56.2
ak_diket_081	diket2	-668.25632	56.4
ak_lac3_132	lac2	-668.2560378	57.2
ak_lac4_023	lac5	-668.2559199	57.5
ak_diket_041	diket2	-668.2552196	59.3
ak_diket_014	diket2	-668.2546777	60.7
ak_lac3_002	lac3	-668.2536419	63.5
ak_diket3_021	diket1	-668.2533345	64.3
ak_diket_055	diket2	-668.2528341	65.6
ak_diket_097	diket2	-668.2523014	67.0
ak_diket_087	diket2	-668.2515845	68.9

Table 17 The 60 lowest energy AO^+ conformers out of 383 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
ao_lac2_001	lac2	-628.9806749	0.0
ao_diket2_010	diket2	-628.9785503	5.6
ao_diket3_135	diket2	-628.9785499	5.6
ao_lac3_009	lac2	-628.978362	6.1
ao_lac2_002	lac2	-628.9783619	6.1
ao_diket3_033	diket2	-628.9770828	9.4
ao_diket2_023	diket2	-628.9770823	9.4
ao_lac2_006	lac2	-628.9768285	10.1
ao_diket3_095	diket2	-628.9763434	11.4
ao_diket2_005	diket2	-628.9763365	11.4
ao_lac3_016	lac2	-628.9756233	13.3
ao_lac2_005	lac2	-628.975613	13.3
ao_diket3_005	diket2	-628.9754819	13.6
ao_diket3_359	diket2	-628.9754811	13.6
ao_diket2_003	diket2	-628.9754762	13.6
ao_diket2_002	diket2	-628.9754203	13.8
ao_diket3_345	diket2	-628.9754162	13.8

ao_lac2_013	lac2	-628.9750755	14.7
ao_lac2_003	lac2	-628.9746823	15.7
ao_lac2_008	lac2	-628.9732073	19.6
ao_lac3_006	lac2	-628.9732021	19.6
ao_lac2_010	lac2	-628.9727586	20.8
ao_lac2_009	lac2	-628.9717177	23.5
ao_lac2_011	lac2	-628.9618669	49.4
ao_lac2_012	lac2	-628.9609135	51.9
ao_diket1_305	diket4	-628.9606707	52.5
ao_diket1_139	diket1	-628.9586983	57.7
ao_diket1_173	diket4	-628.9576312	60.5
ao_diket1_066	diket4	-628.9567209	62.9
ao_diket2_009	diket2	-628.9563977	63.7
ao_diket1_285	diket4	-628.9561278	64.4
ao_diket1_220	diket1	-628.9557302	65.5
ao_diket1_038	diket1	-628.9555055	66.1
ao_diket1_321	diket4	-628.9555014	66.1
ao_lac3_002	lac3	-628.9551428	67.0
ao_lac1_005	lac1	-628.9549317	67.6
ao_oxa3_024	lac1	-628.9549308	67.6
ao_oxa1_067	lac1	-628.9549283	67.6
ao_diket1_086	diket4	-628.9547958	67.9
ao_diket1_085	diket4	-628.9546097	68.4
ao_diket1_284	diket1	-628.9544305	68.9
ao_diket1_159	diket1	-628.9543793	69.0
ao_diket1_256	diket2	-628.954129	69.7
ao_diket1_104	diket1	-628.9539689	70.1
ao_diket1_237	diket1	-628.9539445	70.2
ao_diket1_224	diket1	-628.953899	70.3
ao_lac3_018	lac3	-628.9535447	71.2
ao_diket1_239	diket1	-628.9535231	71.3
ao_diket1_310	diket1	-628.953362	71.7
ao_lac3_023	lac3	-628.953218	72.1
ao_diket2_015	diket2	-628.9529231	72.9
ao_diket1_029	diket2	-628.9528439	73.1
ao_diket2_022	diket2	-628.9527801	73.2
ao_oxa2_001	oxa2	-628.9518923	75.6
ao_oxa3_011	oxa2	-628.9518916	75.6
ao_oxa1_090	oxa2	-628.9518913	75.6
ao_diket2_027	diket2	-628.9512255	77.3
ao_oxa1_172	lac1	-628.951183	77.4
ao_lac1_008	lac1	-628.9511791	77.4
ao_oxa3_052	lac1	-628.9511784	77.4

Table 18 The 60 lowest energy AB^+ conformers out of 136 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Conformer	Conformer	ΔG	ΔΔG
ab_lac1_001	lac4	-589.6818299	0.0
ab_oxa1_032	lac4	-589.6818294	0.0
ab_oxa3_002	lac4	-589.6818264	0.0
ab_lac2_007	lac2	-589.679969	4.9
ab_lac1_003	lac2	-589.6799653	4.9
ab_lac3_003	lac2	-589.6799499	4.9
ab_diket3_009	diket2	-589.6780628	9.9
ab_diket2_001	diket2	-589.6780594	9.9
ab_lac2_008	lac2	-589.6753492	17.0
ab_diket3_062	diket2	-589.6742212	20.0
ab_diket2_005	diket2	-589.6742197	20.0
ab_lac2_004	lac2	-589.6736374	21.5
ab_lac3_001	lac2	-589.6736307	21.5
ab_lac2_009	lac2	-589.6734551	22.0
ab_lac2_003	lac2	-589.6724164	24.7
ab_diket3_002	diket2	-589.6708892	28.7
ab_diket2_003	diket2	-589.6708786	28.8
ab_lac2_005	lac2	-589.6651209	43.9
ab_diket1_071	diket4	-589.65817	62.1
ab_diket1_119	diket4	-589.6572804	64.5
ab_diket1_115	diket4	-589.6569473	65.3
ab_diket1_112	diket4	-589.6551763	70.0
ab_diket1_084	diket1	-589.6534834	74.4
ab_diket1_090	diket1	-589.6534608	74.5
ab_lac3_007	lac3	-589.6495093	84.9
ab_lac3_004	lac3	-589.6490923	86.0
ab_oxa1_004	lac1	-589.6482797	88.1
ab_diket3_130	diket3	-589.647566	90.0
ab_diket3_122	diket3	-589.6469751	91.5
ab_lac3_006	lac3	-589.6451442	96.3
ab_diket3_126	diket3	-589.6443975	98.3
ab_oxa3_003	oxa2	-589.6437169	100.1
ab_oxa2_001	oxa2	-589.643712	100.1
ab_oxa1_001	oxa2	-589.6437104	100.1
ab_diket1_025	diket1	-589.6433859	100.9
ab_diket1_050	diket1	-589.6431921	101.4
ab_diket1_078	diket1	-589.6430598	101.8
ab_oxa1_030	oxa1	-589.6422297	104.0
ab_oxa1_036	oxa1	-589.6422252	104.0

ab_diket3_100	diket3	-589.6419256	104.8
ab_diket3_119	diket3	-589.6419187	104.8
ab_oxa1_042	oxa1	-589.6416419	105.5
ab_diket1_065	diket1	-589.6405459	108.4
ab_oxa1_044	oxa2	-589.6404466	108.7
ab_oxa2_003	oxa2	-589.6404459	108.7
ab_diket1_082	diket1	-589.6403351	108.9
ab_oxa2_002	oxa2	-589.6400771	109.6
ab_oxa3_001	oxa2	-589.6400733	109.6
ab_diket3_033	diket3	-589.6396206	110.8
ab_lac1_002	lac1	-589.6396039	110.9
ab_diket3_112	diket3	-589.6395425	111.0
ab_diket3_125	diket3	-589.6395362	111.0
ab_diket1_083	diket1	-589.6394411	111.3
ab_oxa2_006	oxa2	-589.6393543	111.5
ab_oxa2_005	oxa2	-589.639224	111.9
ab_oxa1_045	oxa2	-589.6392175	111.9
ab_diket3_056	diket3	-589.639179	112.0
ab_oxa1_051	lac1	-589.6390784	112.2
ab_diket3_017	diket3	-589.6390006	112.4
ab_diket3_073	diket3	-589.638337	114.2

Table 19 All 42 optimized AZ^+ structures. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
az_diket2_001	diket2	-550.3670408	0.0
az_diket3_005	diket2	-550.3670391	0.0
az_diket2_002	diket2	-550.3625632	11.8
az_diket3_002	diket2	-550.3625621	11.8
az_diket1_018	diket1	-550.3582133	23.2
az_diket1_024	diket1	-550.3504042	43.7
az_diket1_014	diket1	-550.3462351	54.6
az_lac2_001	lac2	-550.3455517	56.4
az_diket3_013	diket3	-550.3448976	58.1
az_diket3_023	diket3	-550.3438915	60.8
az_diket1_012	diket1	-550.3431494	62.7
az_diket3_036	diket3	-550.3430197	63.1
az_oxa1_017	oxa1	-550.3422206	65.2
az_diket3_018	diket3	-550.3417065	66.5
az_oxa1_009	oxa1	-550.3415463	66.9
az_diket1_022	diket1	-550.339224	73.0
az_diket1_007	diket1	-550.3381857	75.8

az_diket3_039	diket3	-550.3377495	76.9
az_diket3_038	diket3	-550.3362661	80.8
az_oxa1_005	oxa1	-550.3357705	82.1
az_oxa1_019	oxa1	-550.3331471	89.0
az_lac3_001	lac2	-550.3328058	89.9
az_oxa1_021	oxa1	-550.3326826	90.2
az_oxa1_020	oxa1	-550.3325814	90.5
az_oxa3_009	oxa3	-550.3322373	91.4
az_oxa3_006	oxa3	-550.3279802	102.6
az_oxa3_004	oxa3	-550.3269501	105.3
az_oxa2_007	oxa2	-550.3267376	105.8
az_oxa2_016	oxa2	-550.3267311	105.8
az_oxa1_024	oxa1	-550.3264914	106.5
az_oxa1_026	oxa1	-550.3259063	108.0
az_oxa2_003	oxa2	-550.3255735	108.9
az_oxa2_006	oxa2	-550.3254584	109.2
az_oxa2_001	oxa2	-550.3253349	109.5
az_oxa2_014	oxa2	-550.3251092	110.1
az_oxa2_018	oxa2	-550.3250885	110.1
az_oxa2_015	oxa2	-550.3250265	110.3
az_lac1_001	lac1	-550.3248632	110.7
az_oxa2_010	oxa2	-550.3248485	110.8
az_oxa2_013	oxa2	-550.3241236	112.7
az_oxa2_011	oxa2	-550.323699	113.8

Table 20 The 60 lowest energy AAK^+ conformers out of 1865 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
aak_lac3_123	lac2	-915.6127729	0.0
aak_lac4_008	lac2	-915.6127729	0.0
aak_lac6_123	lac2	-915.6127729	0.0
aak_lac2_079	lac2	-915.6127698	0.0
aak_lac5_154	lac3	-915.6123243	1.2
aak_lac5_039	lac3	-915.6117215	2.8
aak_lac2_019	lac2	-915.6083369	11.6
aak_lac7_109	lac2	-915.6083287	11.7
aak_lac2_004	lac2	-915.6081896	12.0
aak_lac5_008	lac2	-915.6070776	15.0
aak_lac2_015	lac2	-915.6070715	15.0
aak_lac5_202	lac5	-915.6068542	15.5
aak_lac5_063	lac3	-915.6065818	16.3
aak_lac2_007	lac2	-915.6063843	16.8

aak_lac3_172	lac2	-915.606383	16.8
aak_lac6_172	lac2	-915.606383	16.8
aak_lac4_083	lac2	-915.606382	16.8
aak_lac7_083	lac2	-915.606382	16.8
aak_lac5_129	lac3	-915.606255	17.1
aak_lac2_018	lac2	-915.6057184	18.5
aak_lac5_180	lac3	-915.6056803	18.6
aak_lac5_108	lac2	-915.6055863	18.9
aak_lac2_084	lac2	-915.6055798	18.9
aak_lac5_035	lac2	-915.6051117	20.1
aak_lac2_005	lac2	-915.6051101	20.1
aak_lac2_001	lac2	-915.6050595	20.3
aak_lac5_133	lac3	-915.6048794	20.7
aak_lac3_195	lac2	-915.604593	21.5
aak_lac6_195	lac2	-915.604593	21.5
aak_lac5_118	lac3	-915.6040322	22.9
aak_lac5_042	lac3	-915.6034478	24.5
aak_lac5_004	lac3	-915.6034259	24.5
aak_lac2_031	lac2	-915.6030921	25.4
aak_lac2_017	lac2	-915.6028198	26.1
aak_lac4_190	lac2	-915.602799	26.2
aak_lac7_190	lac2	-915.602799	26.2
aak_lac2_016	lac2	-915.6026066	26.7
aak_lac5_153	lac2	-915.6026052	26.7
aak_lac5_047	lac3	-915.6024517	27.1
aak_lac2_012	lac2	-915.6021145	28.0
aak_lac5_086	lac2	-915.60211	28.0
aak_lac5_062	lac2	-915.6019656	28.4
aak_lac2_090	lac2	-915.6019583	28.4
aak_lac2_013	lac2	-915.6016997	29.1
aak_lac2_014	lac2	-915.6014235	29.8
aak_lac5_160	lac3	-915.6012343	30.3
aak_lac5_010	lac2	-915.6011894	30.4
aak_lac2_043	lac2	-915.6011843	30.4
aak_lac5_113	lac4	-915.6009066	31.2
aak_lac2_034	lac2	-915.6008678	31.3
aak_lac5_040	lac2	-915.6008534	31.3
aak_lac5_016	lac4	-915.6008375	31.3
aak_lac2_058	lac2	-915.6006693	31.8
aak_lac5_023	lac4	-915.6006074	31.9
aak_lac2_022	lac2	-915.6004858	32.3
aak_lac2_081	lac2	-915.6004644	32.3
aak_lac5_038	lac3	-915.6004571	32.3

aak_lac5_156	lac2	-915.6004174	32.4
aak_lac2_023	lac2	-915.6004147	32.4
aak_lac4_011	lac2	-915.6003136	32.7

Table 21 The 60 lowest energy AAO^+ conformers out of 576 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
aao_lac2_001	lac2	-876.3186973	0.0
aao_lac5_027	lac4	-876.3175436	3.0
aao_lac5_043	lac4	-876.3166901	5.3
aao_lac2_011	lac2	-876.3164198	6.0
aao_lac3_019	lac2	-876.3161603	6.7
aao_lac2_003	lac2	-876.316155	6.7
aao_lac1_041	lac2	-876.3149519	9.8
aao_lac2_004	lac2	-876.3149503	9.8
aao_lac3_084	lac3	-876.3145976	10.8
aao_lac3_037	lac3	-876.3138297	12.8
aao_lac3_083	lac3	-876.3134338	13.8
aao_lac4_004	lac3	-876.3134318	13.8
aao_lac3_051	lac4	-876.3132087	14.4
aao_lac2_010	lac2	-876.3129852	15.0
aao_lac2_005	lac2	-876.3129764	15.0
aao_lac3_048	lac2	-876.3129699	15.0
aao_lac3_058	lac3	-876.3127499	15.6
aao_lac4_001	lac3	-876.3127395	15.6
aao_lac5_012	lac5	-876.312686	15.8
aao_lac5_025	lac5	-876.3125875	16.0
aao_lac3_088	lac3	-876.3124637	16.4
aao_lac2_002	lac2	-876.3120072	17.6
aao_lac3_012	lac2	-876.3118288	18.0
aao_lac2_015	lac2	-876.3118267	18.0
aao_lac3_082	lac4	-876.3115355	18.8
aao_lac2_024	lac2	-876.3110999	19.9
aao_lac4_006	lac3	-876.3110656	20.0
aao_lac3_086	lac3	-876.3107937	20.8
aao_lac4_020	lac3	-876.3107907	20.8
aao_lac5_020	lac5	-876.3107539	20.9
aao_lac3_075	lac4	-876.3101558	22.4
aao_lac3_025	lac2	-876.3099009	23.1
aao_lac2_030	lac2	-876.309899	23.1
aao_lac5_029	lac2	-876.3096619	23.7
aao_lac3_079	lac3	-876.3096424	23.8

aao_lac3_013	lac3	-876.3093899	24.4
aao_lac4_005	lac4	-876.3092918	24.7
aao_lac5_010	lac5	-876.3092456	24.8
aao_lac2_008	lac2	-876.3089231	25.7
aao_lac3_072	lac2	-876.308871	25.8
aao_lac4_024	lac3	-876.3087537	26.1
aao_lac3_092	lac3	-876.3087511	26.1
aao_lac4_009	lac3	-876.3086784	26.3
aao_lac3_066	lac3	-876.3086701	26.3
aao_lac3_002	lac3	-876.3085522	26.6
aao_lac2_018	lac2	-876.3085001	26.8
aao_lac2_007	lac2	-876.307984	28.1
aao_lac3_040	lac4	-876.3078835	28.4
aao_lac5_008	lac5	-876.3077416	28.8
aao_lac2_013	lac2	-876.3075317	29.3
aao_lac3_006	lac2	-876.3075138	29.4
aao_lac4_010	lac4	-876.3073318	29.8
aao_lac5_013	lac4	-876.3072124	30.2
aao_lac2_025	lac2	-876.3070362	30.6
aao_lac2_033	lac2	-876.3069551	30.8
aao_lac5_018	lac2	-876.306945	30.9
aao_lac3_036	lac3	-876.3069344	30.9
aao_lac4_021	lac4	-876.3069268	30.9
aao_lac5_037	lac5	-876.3068745	31.0
aao_lac4_014	lac5	-876.3068624	31.1

Table 22 The 60 lowest energy AAB^+ conformers out of 440 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Conformer	Structure	ΔG	ΔΔG
aab_lac1_0050	lac2	-837.0185456	0.0
aab_lac2_0004	lac2	-837.0185411	0.0
aab_lac3_0029	lac2	-837.016862	4.4
aab_lac1_0020	lac2	-837.016859	4.4
aab_lac2_0003	lac2	-837.0168543	4.4
aab_lac2_0005	lac2	-837.0164292	5.6
aab_lac3_0108	lac2	-837.0143181	11.1
aab_lac4_0101	lac4	-837.0140711	11.7
aab_lac5_0013	lac4	-837.0140679	11.8
aab_lac1_0026	lac2	-837.0139758	12.0
aab_lac3_0100	lac2	-837.0139654	12.0
aab_lac3_0099	lac3	-837.0132205	14.0
aab_lac4_0095	lac3	-837.0132174	14.0

aab_lac5_0049	lac2	-837.0126891	15.4
aab_lac4_0059	lac2	-837.0126781	15.4
aab_lac2_0002	lac2	-837.0126768	15.4
aab_lac1_0067	lac2	-837.0126755	15.4
aab_lac3_0053	lac3	-837.0123892	16.2
aab_lac1_0005	lac4	-837.0123036	16.4
aab_oxa3_271	lac4	-837.0123012	16.4
aab_oxa5_007	lac4	-837.0122979	16.4
aab_lac4_0056	lac3	-837.0122973	16.4
aab_lac5_0064	lac3	-837.0116907	18.0
aab_lac3_0070	lac3	-837.011687	18.0
aab_lac3_0039	lac5	-837.0113794	18.8
aab_lac4_0019	lac3	-837.0113144	19.0
aab_lac3_0060	lac3	-837.0113133	19.0
aab_lac3_0041	lac3	-837.0108366	20.2
aab_lac4_0002	lac3	-837.0108198	20.3
aab_lac3_0058	lac3	-837.0105584	21.0
aab_lac4_0068	lac3	-837.0105478	21.0
aab_lac4_0110	lac2	-837.0098873	22.7
aab_lac3_0078	lac5	-837.0095745	23.6
aab_lac5_0004	lac5	-837.0095533	23.6
aab_lac1_0018	lac2	-837.0090641	24.9
aab_lac2_0006	lac2	-837.0090605	24.9
aab_lac3_0031	lac3	-837.0090543	24.9
aab_lac4_0086	lac3	-837.0090483	24.9
aab_lac3_0097	lac2	-837.0088243	25.5
aab_lac3_0067	lac3	-837.0087118	25.8
aab_lac4_0107	lac3	-837.0087015	25.8
aab_lac2_0012	lac2	-837.0079879	27.7
aab_lac2_0009	lac2	-837.007929	27.9
aab_lac3_0006	lac5	-837.0078932	28.0
aab_lac5_0029	lac5	-837.0078919	28.0
aab_lac2_0016	lac2	-837.0078888	28.0
aab_lac4_0041	lac4	-837.007843	28.1
aab_lac3_0004	lac2	-837.0076899	28.5
aab_lac2_0015	lac2	-837.0076812	28.5
aab_lac2_0014	lac2	-837.007569	28.8
aab_oxa3_512	lac2	-837.0074648	29.1
aab_lac2_0007	lac2	-837.0073313	29.4
aab_lac4_0001	lac4	-837.0071062	30.0
aab_lac5_0007	lac4	-837.0071051	30.0
aab_lac4_0115	lac2	-837.006686	31.1
aab_lac5_0001	lac5	-837.0065752	31.4

aab_lac3_0009	lac5	-837.0065747	31.4
aab_lac3_0008	lac3	-837.0065514	31.5
aab_lac3_0104	lac3	-837.0062241	32.3
aab_lac5_0020	lac4	-837.005792	33.5

Table 23 The 60 lowest energy AAZ^+ conformers out of 346 conformers optimized. ΔG energies are given in Hartrees while ΔΔG is reported in kJ/mol. All energies are reported at 298K.

Name	Conformer	ΔG	ΔΔG
aaz_lac2_0002	lac2	-797.6865439	0.0
aaz_lac3_0042	lac2	-797.6865438	0.0
aaz_lac3_0069	lac2	-797.6862363	0.8
aaz_lac2_0010	lac2	-797.6855674	2.6
aaz_lac2_0003	lac2	-797.6849128	4.3
aaz_lac5_0038	lac2	-797.6849077	4.3
aaz_lac3_0101	lac2	-797.6834148	8.2
aaz_lac2_0007	lac2	-797.6831022	9.0
aaz_lac2_0001	lac2	-797.6822075	11.4
aaz_lac2_0005	lac2	-797.6816197	12.9
aaz_lac4_0049	lac2	-797.6816134	12.9
aaz_lac2_0025	lac2	-797.6809567	14.7
aaz_lac3_0015	lac2	-797.6809539	14.7
aaz_macro4_0305	macro1	-797.6806928	15.4
aaz_macro2_1580	macro1	-797.6806843	15.4
aaz_macro1_0001	macro1	-797.6806807	15.4
aaz_lac5_0020	lac2	-797.6806267	15.5
aaz_lac3_0109	lac2	-797.680623	15.5
aaz_lac3_0040	lac3	-797.6803864	16.2
aaz_lac3_0084	lac3	-797.6802665	16.5
aaz_lac4_0008	lac3	-797.6802606	16.5
aaz_macro2_0772	macro1	-797.6801692	16.7
aaz_macro1_0002	macro1	-797.6801691	16.7
aaz_macro4_0207	macro1	-797.6801689	16.7
aaz_lac2_0008	lac2	-797.6799654	17.3
aaz_lac3_0059	lac2	-797.6799636	17.3
aaz_lac2_0031	lac2	-797.6799235	17.4
aaz_lac2_0023	lac2	-797.6795558	18.3
aaz_lac3_0020	lac2	-797.6795389	18.4
aaz_oxa1_0021	oxa1	-797.6792261	19.2
aaz_oxa4_0007	oxa1	-797.6792252	19.2
aaz_oxa3_1198	oxa1	-797.679223	19.2
aaz_lac4_0001	lac3	-797.6791871	19.3
aaz_lac4_0057	lac3	-797.6791444	19.4

aaz_lac3_0029	lac3	-797.6791373	19.4
aaz_lac2_0046	lac2	-797.6785521	21.0
aaz_lac5_0053	lac3	-797.6785513	21.0
aaz_lac3_0053	lac3	-797.6785492	21.0
aaz_oxa2_0240	oxa1	-797.6784196	21.3
aaz_oxa1_0004	oxa1	-797.6784191	21.3
aaz_oxa3_0440	oxa1	-797.678416	21.3
aaz_oxa4_0004	oxa1	-797.6784106	21.4
aaz_lac3_0046	lac3	-797.6783926	21.4
aaz_lac3_0022	lac3	-797.6783484	21.5
aaz_lac2_0042	lac2	-797.67805	22.3
aaz_lac5_0008	lac2	-797.6780456	22.3
aaz_lac3_0036	lac3	-797.6778772	22.8
aaz_lac2_0049	lac2	-797.6778253	22.9
aaz_lac3_0065	lac2	-797.6778216	22.9
aaz_oxa1_0002	oxa1	-797.6775356	23.7
aaz_oxa4_0002	oxa1	-797.6775356	23.7
aaz_lac3_0018	lac2	-797.6773909	24.0
aaz_lac2_0019	lac2	-797.6773896	24.0
aaz_lac3_0047	lac3	-797.677016	25.0
aaz_macro2_1530	macro1	-797.6768247	25.5
aaz_macro1_0008	macro1	-797.6768186	25.5
aaz_lac5_0010	lac4	-797.6767867	25.6
aaz_lac4_0045	lac4	-797.6767683	25.7
aaz_lac4_0016	lac2	-797.6764968	26.4
aaz_lac2_0015	lac2	-797.6764626	26.5

Figures

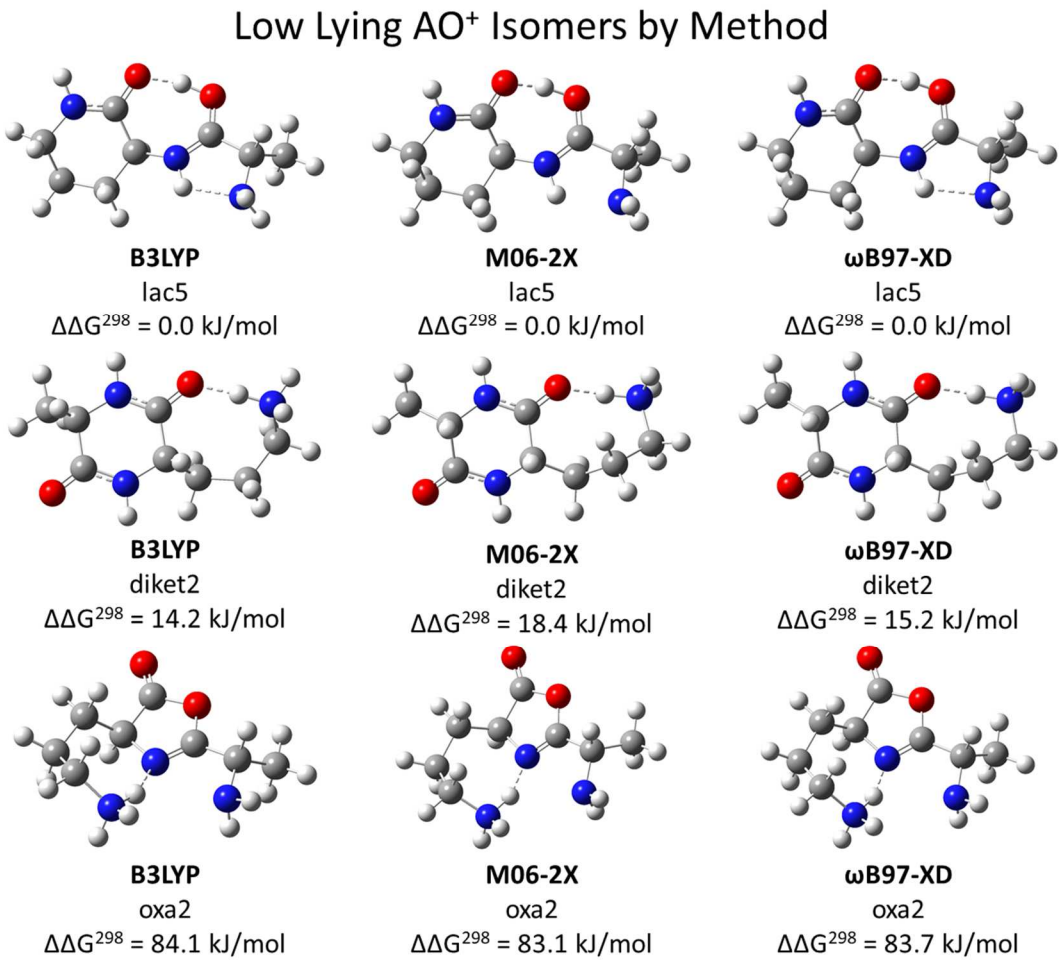


Figure 39 Lowest lying AO^+ structure for each isomer type as computed by the B3LYP (Left), M06-2X (Middle), and ωB97-XD (Right) methods.

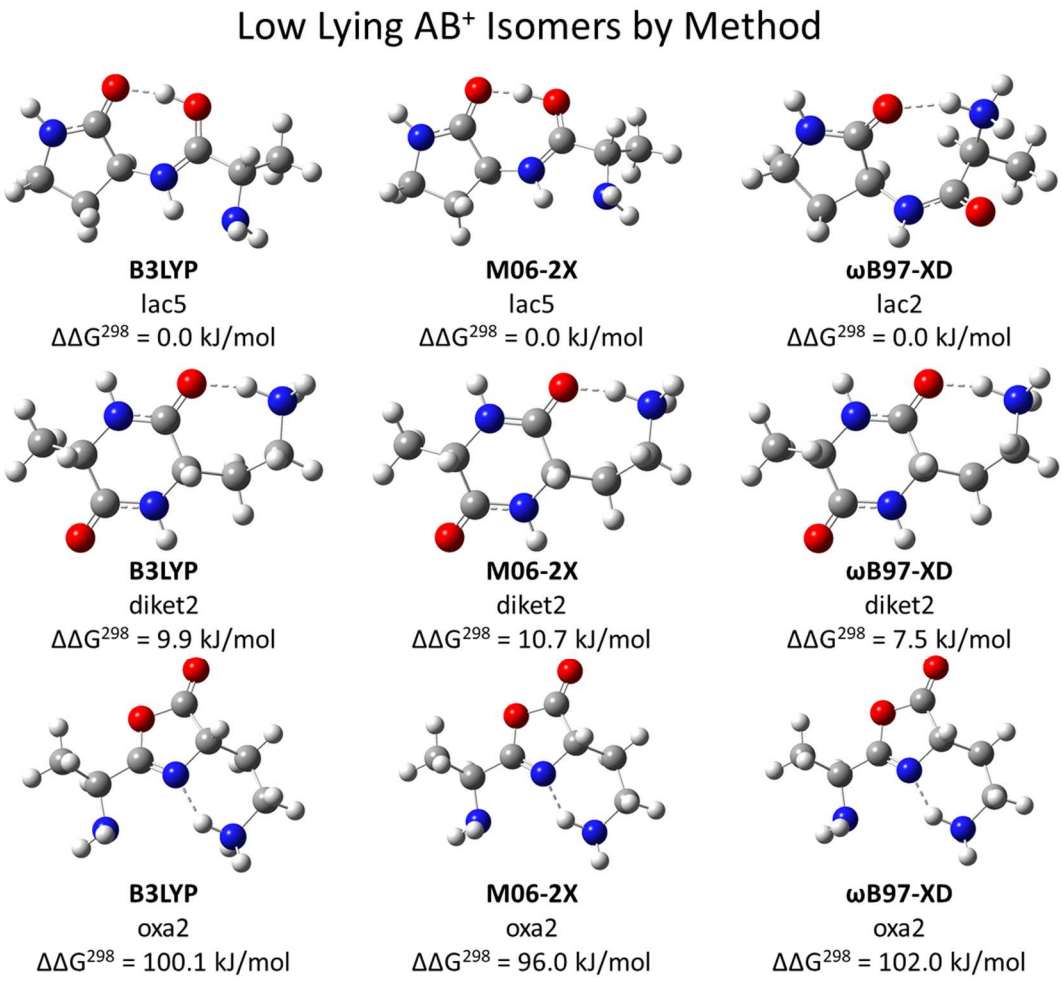


Figure 40 Lowest lying AB^+ structure for each isomer type as computed by the B3LYP (Left), M06-2X (Middle), and ωB97-XD (Right) methods.

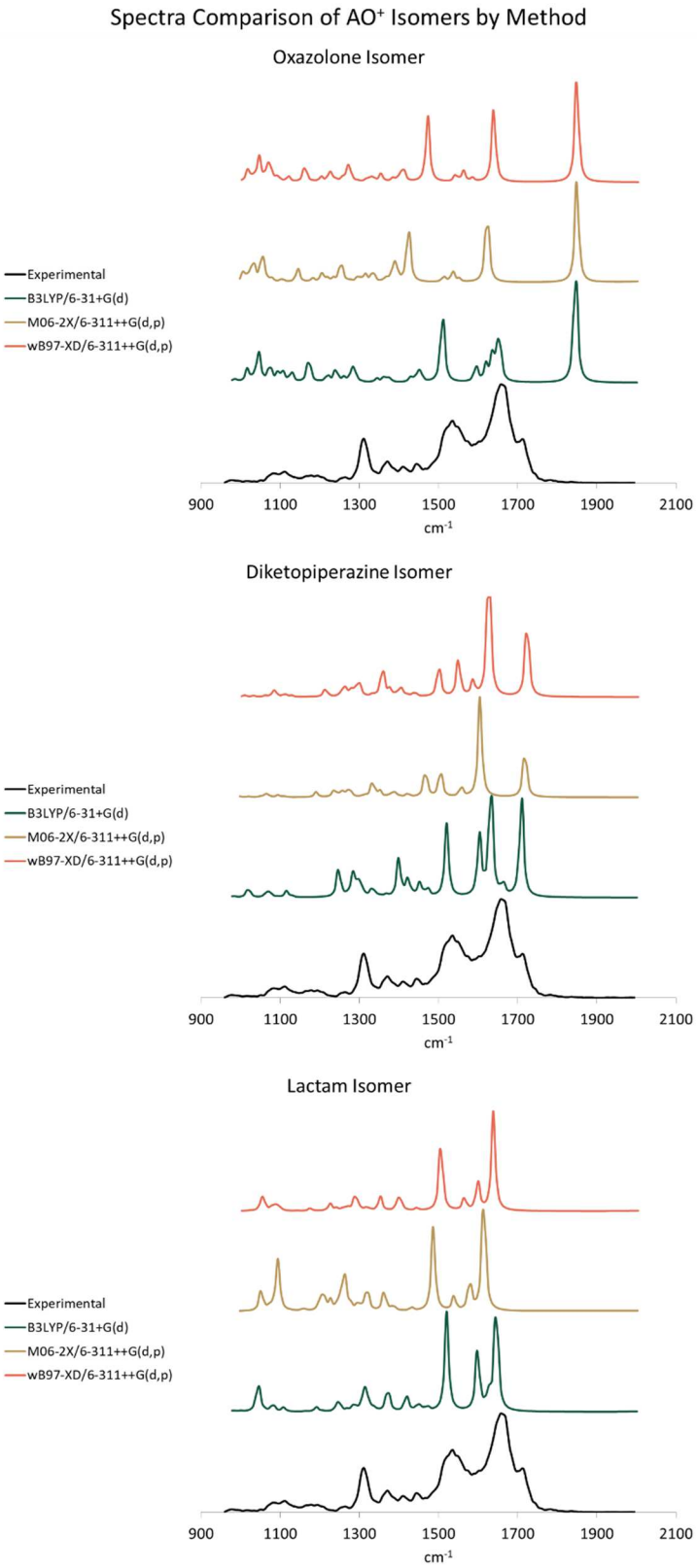


Figure 41 Calculated spectra for the AO^+ Oxazolone (Top), Diketopiperazine (Middle), and Lactam (Bottom) isomers as compared to the experimental spectrum collected from FELIX.

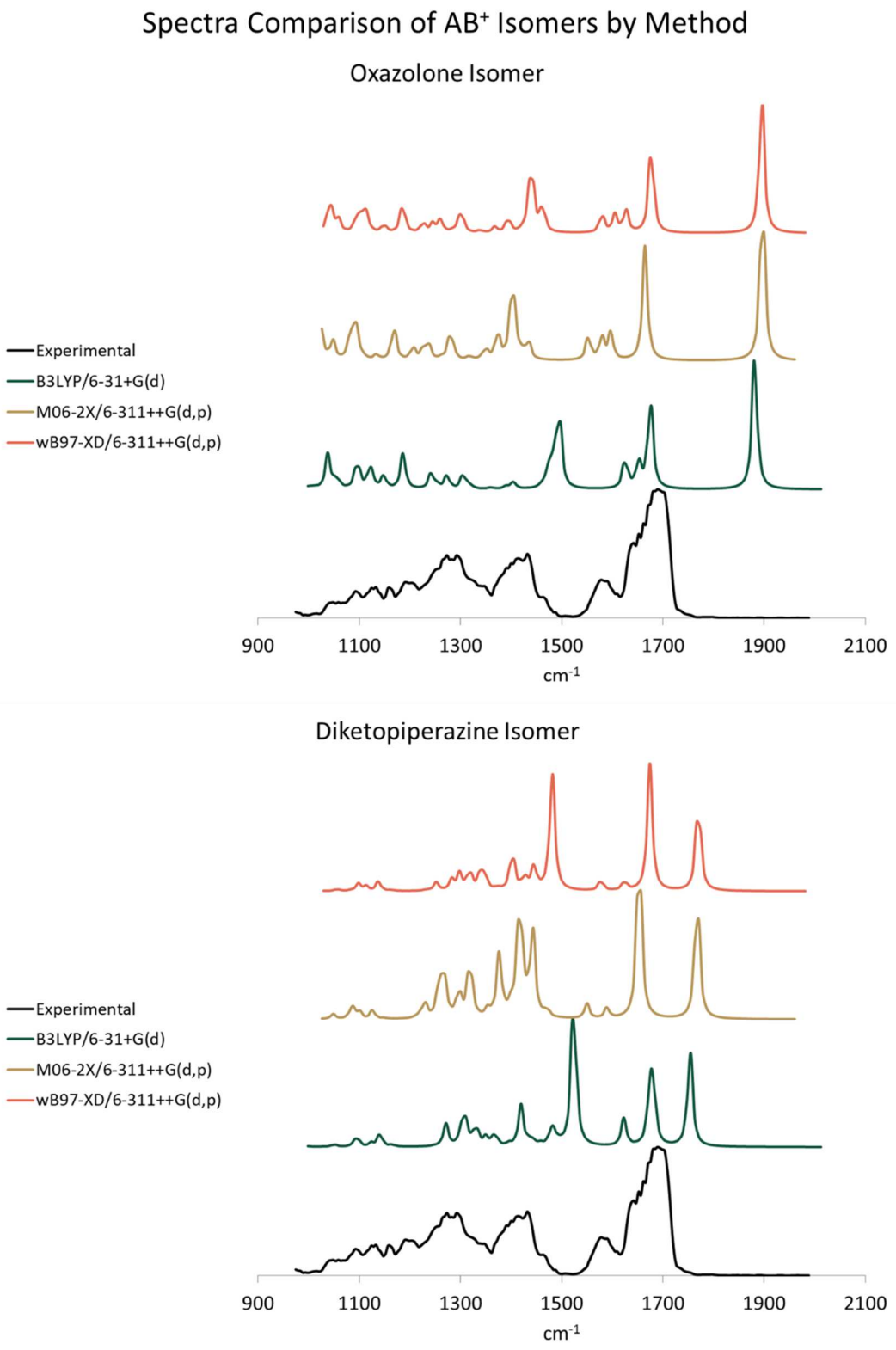


Figure 42 Calculated spectra for the AB^+ Oxazolone (Top) and Diketopiperazine (Bottom) isomers as compared to the experimental spectrum collected from CLIO.

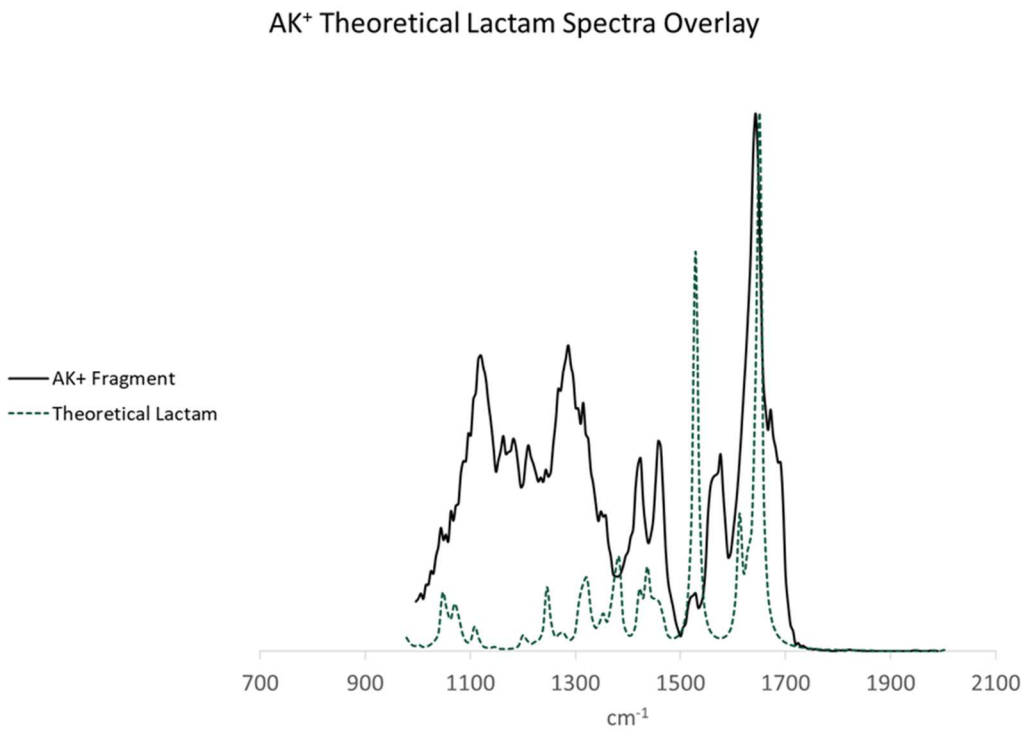


Figure 43 An overlay comparison of the lowest energy AK^+ theoretical lactam spectrum (Dashed Green) and the experimental data gathered from CLIO (Solid Black).

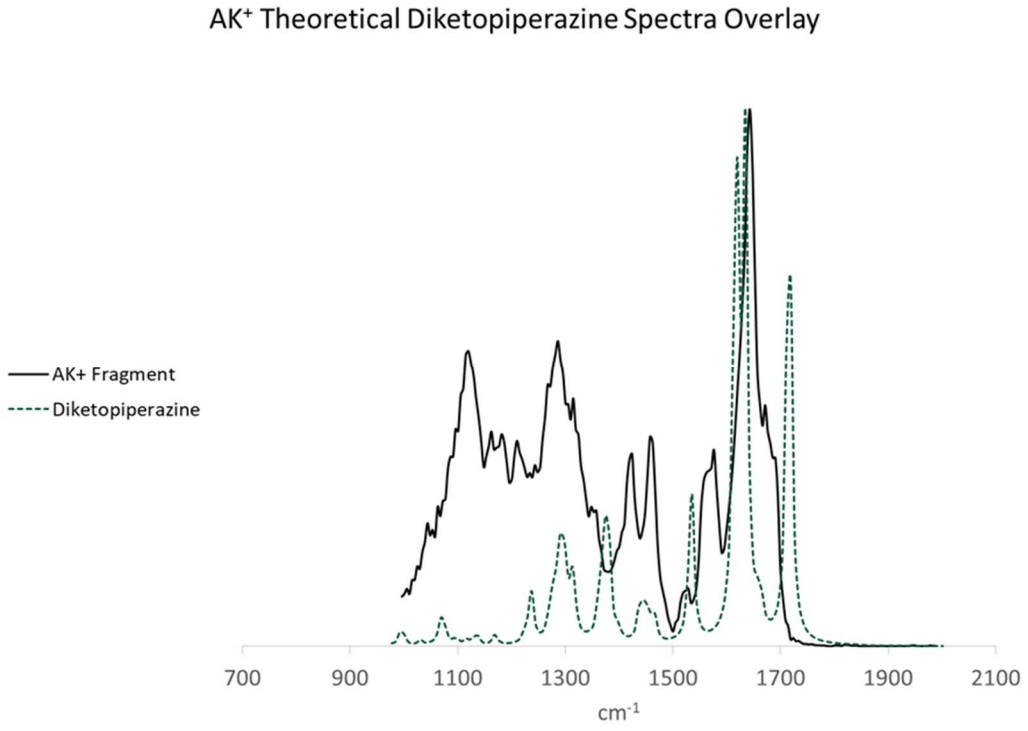


Figure 44 An overlay comparison of the lowest energy AK^+ theoretical diketopiperazine spectrum (Dashed Green) and the experimental data gathered from CLIO (Solid Black).

Bibliography

(1) Cohen, S. L.; Chait, B. T. Influence of Matrix Solution Conditions on the MALDI-MS Analysis of Peptides and Proteins. *Anal. Chem.* **1996**, *68* (1), 31–37.

(2) Caprioli, R. M.; Farmer, T. B.; Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* **1997**, *69* (23), 4751–4760.

(3) Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. The Characteristics of Peptide Collision-Induced Dissociation Using a High-Performance MALDI-TOF/TOF Tandem Mass Spectrometer. *Anal. Chem.* **2000**, *72* (3), 552–558.

(4) Loo, J. A. Electrospray Ionization Mass Spectrometry: A Technology for Studying Noncovalent Macromolecular Complexes. *Int. J. Mass Spectrom.* **2000**, *200* (1–3), 175–186.

(5) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal. Chem.* **2004**, *76* (14), 3935–3943.

(6) Krückeberg, S.; Schooss, D.; Maier-Borst, M.; Parks, J. H. Diffraction of Trapped (CsI)NCs+: The Appearance of Bulk Structure. *Phys. Rev. Lett.* **2000**, *85* (21), 4494–4497.

(7) Tverdova, N. V.; Pogonin, A. E.; Ischenko, A. A.; Rumyantseva, V. D.; Koifman, O. I.; Giricheva, N. I.; Girichev, G. V. Combined Gas-Phase Electron Diffraction/Mass Spectrometry and DFT Study of the Molecular Structure of Zinc(II) Etioporphyrin-II. *Struct. Chem.* **2015**, *26* (5–6), 1521–1530.

(8) Shlykov, S. A.; Phien, T. D.; Trang, N. H. Orbital Interaction between Electron Lone Pair and Carbonyl Group in N-Trifluoroacetylpiperidine and N-Piperidine Amides: Planar and Non-Planar Nitrogen Bond Configurations. *Tetrahedron* **2017**, *73* (35), 5311–5320.

(9) Dakkouri, M.; Girichev, G.; Giricheva, N.; Petrov, V.; Petrova, V. Structural Analysis and Probing the Conformational Space of Dansylamide by Means of Gas-Phase Electron Diffraction and Quantum Chemistry. *Struct. Chem.* **2018**, *29* (3), 823–835.

(10) Rosenberg, J.; Parker, W. R.; Cammarata, M. B.; Brodbelt, J. S. UV-POSIT: Web-Based Tools for Rapid and Facile Structural Interpretation of Ultraviolet Photodissociation (UVPD) Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (6), 1323–1326.

(11) Greer, S. M.; Brodbelt, J. S. Top-Down Characterization of Heavily Modified Histones Using 193 Nm Ultraviolet Photodissociation Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (3), 1138–1145.

(12) Dwivedi, P.; Wu, C.; Matz, L. M.; Clowers, B. H.; Siems, W. F.; Hill Jr., H. H. Gas-Phase Chiral Separations by Ion Mobility Spectrometry. *Anal. Chem.* **2006**, *78* (24), 8200–8206.

(13) Yehia Ibrahim; Edreese Alsharaeh; Ridha Mabrouki; Paul Momoh; Enli Xie, and; El-Shall*, M. S.; Ibrahim, Y.; Alsharaeh, E.; Mabrouki, R.; Momoh, P.; et al. Ion Mobility of Ground and Excited States of Laser-Generated Transition Metal Cations. *J. Phys. Chem. A* **2008**, *112* (6), 1112–1124.

(14) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.; Robinson, C. V. Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3* (7), 1139–1152.

(15) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557–9565.

(16) Schwarz, H. A. Gas Phase Infrared Spectra of Oxonium Hydrate Ions from 2 to 5 μ. *J. Chem. Phys.* **1977**, *67* (12), 5525–5534.

(17) Bagratashvili, V. N.; Letokhov, V. S.; Makarov, A. A.; Ryabov, E. A. Vol.4: Multiple Photon Infrared Laser Photophysics and Photochemistry. IV. *Laser Chem.* **1984**, *4* (5), 311–423.

(18) Ashfold, M. N. R.; Hancock, G.; Ketley, G. Infrared Multiple Photon Excitation and Dissociation of Simple Molecules. *Phys. Org. Chem.* **1979**, *67*, 204–211.

(19) Polfer, N. C. Infrared Multiple Photon Dissociation Spectroscopy of Trapped Ions. *Chemical Society Reviews*. The Royal Society of Chemistry April 18, 2011, pp 2211–2221.

(20) Dunbar, R. C.; McMahon, T. B. *Activation of Unimolecular Reactions by Ambient Blackbody Radiation*; American Association for the Advancement of Science, 1998; Vol. 279, pp 194–197.

(21) Jeanne Dit Fouque, K.; Lavanant, H.; Zirah, S.; Steinmetz, V.; Rebuffat, S.; Maître, P.; Afonso, C. IRMPD Spectroscopy: Evidence of Hydrogen Bonding in the Gas Phase Conformations of Lasso Peptides and Their Branched-Cyclic Topoisomers. *J. Phys. Chem. A* **2016**, *120* (21), 3810–3816.

(22) van Stipdonk, M. J.; Patterson, K.; Gibson, J. K.; Berden, G.; Oomens, J. IRMPD Spectroscopy Reveals a Novel Rearrangement Reaction for Modified Peptides That Involves Elimination of the N-Terminal Amino Acid. *Int. J. Mass Spectrom.* **2015**, *379*, 165–178.

(23) Fukui, K.; Takada, Y.; Sumiyoshi, T.; Imai, T.; Takahashi, K. Infrared Multiphoton Dissociation Spectroscopic Analysis of Peptides and Oligosaccharides by Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with a Midinfrared Free-Electron Laser. *J. Phys. Chem. B* **2006**, *110* (32), 16111–16116.

(24) Prell, J. S.; Flick, T. G.; Oomens, J.; Berden, G.; Williams, E. R. Coordination of Trivalent Metal Cations to Peptides: Results from IRMPD Spectroscopy and Theory. *J. Phys. Chem. A* **2010**, *114* (2), 854–860.

(25) Smith, Z. M.; Steinmetz, V.; Martens, J.; Oomens, J.; Poutsma, J. C. Infrared Multiple Photon Dissociation Spectroscopy of Cationized Canavanine: Side-Chain Substitution Influences Gas-Phase Zwitterion Formation. *International Journal of Mass Spectrometry*. Elsevier June 1, 2017, pp 158–173.

(26) Wu, R.; McMahon, T. B. An Investigation of Protonation Sites and Conformations of Protonated Amino Acids by IRMPD Spectroscopy. *ChemPhysChem* **2008**, *9* (18), 2826–2835.

(27) Sinha, R. K.; Erlekam, U.; Bythell, B. J.; Paizs, B.; Maître, P. Diagnosing the Protonation Site of b 2 Peptide Fragment Ions Using IRMPD in the X–H (X = O, N, and C) Stretching Region. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (9), 1645–1650.

(28) Lengyel, J.; Ončák, M.; Herburger, A.; van der Linde, C.; Beyer, M. K. Infrared Spectroscopy of O˙ − and OH − in Water Clusters: Evidence for Fast Interconversion between O˙ − and OH˙OH −. *Phys. Chem. Chem. Phys.* **2017**, *19* (37), 25346–25351.

(29) Ligare, M. R.; Rijs, A. M.; Berden, G.; Kabeláč, M.; Nachtigallova, D.; Oomens, J.; de Vries, M. S. Resonant Infrared Multiple Photon Dissociation Spectroscopy of Anionic Nucleotide Monophosphate Clusters. *J. Phys. Chem. B* **2015**, *119* (25), 7894–7901.

(30) Leavitt, C. M.; Oomens, J.; Dain, R. P.; Steill, J.; Groenewold, G. S.; Stipdonk, M. J. IRMPD Spectroscopy of Anionic Group II Metal Nitrate Cluster Ions. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (5), 772–782.

(31) Polfer, N. C.; Valle, J. J.; Moore, D. T.; Oomens, J.; Eyler, J. R.; Bendiak, B. Differentiation of Isomers by Wavelength-Tunable Infrared Multiple-Photon Dissociation-Mass Spectrometry: Application to Glucose-Containing Disaccharides. *Anal. Chem.* **2006**, *78* (3), 670–679.

(32) Gucinski, A. C.; Somogyi, Á.; Chamot-Rooke, J.; Wysocki, V. H. Separation and Identification of Structural Isomers by Quadrupole Collision-Induced Dissociation-Hydrogen/Deuterium Exchange-Infrared Multiphoton Dissociation (QCID-HDX-IRMPD). *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1329–1338.

(33) Koboldt, D. C.; Steinberg, K. M.; Larson, D. E.; Wilson, R. K.; Mardis, E. R. XThe Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell*. Elsevier September 26, 2013, pp 27–38.

(34) Pandey, A.; Mann, M. Proteomics to Study Genes and Genomes. *Nature*. June 15, 2000, pp 837–846.

(35) Resing, K. A.; Ahn, N. G. Proteomics Strategies for Protein Identification. In *FEBS Letters*; No longer published by Elsevier, 2005; Vol. 579, pp 885–889.

(36) Vyatkina, K.; Dekker, L. J. M.; Wu, S.; VanDuijn, M. M.; Liu, X.; Tolić, N.; Luider, T. M.; Paša-Tolić, L. De Novo Sequencing of Peptides from High-Resolution Bottom-Up Tandem Mass Spectra Using Top-Down Intended Methods. *Proteomics* **2017**, *17* (23–24), 1600321.

(37) Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Anal. Chem.* **2003**, *75* (9), 1963–1971.

(38) Loo, J. A.; Muenster, H. Magnetic Sector-Ion Trap Mass Spectrometry with Electrospray Ionization for High Sensitivity Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **1999**, *13* (1), 54–60.

(39) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of

Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society*. American Chemical Society 1998, pp 3265–3266.

(40) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **1999**, *121* (22), 5142–5154.

(41) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.

(42) Harrison, A. G.; Csizmadia, I. G.; Tang, T. H. Structure and Fragmentation of B2ions in Peptide Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (5), 427–436.

(43) Armentrout, P. B.; Heaton, A. L. Thermodynamics and Mechanisms of Protonated Diglycine Decomposition: A Computational Study. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (4), 621–631.

(44) Bythell, B. J.; Csonka, I. P.; Suhai, S.; Barofsky, D. F.; Paizs, B. Gas-Phase Structure and Fragmentation Pathways of Singly Protonated Peptides with N-Terminal Arginine. *J. Phys. Chem. B* **2010**, *114* (46), 15092–15105.

(45) McGee, W. M.; McLuckey, S. A. The Ornithine Effect in Peptide Cation Dissociation. *J. Mass Spectrom.* **2013**, *48* (7), 856–861.

(46) Crittenden, C. M.; Parker, W. R.; Jenner, Z. B.; Bruns, K. A.; Akin, L. D.; McGee, W. M.; Ciccimaro, E.; Brodbelt, J. S. Exploitation of the Ornithine Effect Enhances Characterization of Stapled and Cyclic Peptides. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (5), 856–863.

(47) Corso, G.; Esposito, M.; Gallo, M.; Russo, A. Dello; Antonio, M. Transformation of Arginine into Ornithine during the Preparation of Its Tert-butyldimethylsilyl Derivative for Analysis by Gas Chromatography/Mass Spectrometry. *Biol. Mass Spectrom.* **1993**, *22* (12), 698–702.

(48) Kuyama, H.; Nakajima, C.; Nakazawa, T.; Nishimura, O. Conversion of Arginine to Ornithine for Improving the Fragmentation Pattern of Peptides Labeled with the N-Terminal Tris(2,4,6-Trimethoxyphenyl)Phosphonium Group in Tandem Mass Spectrometry. *Anal. Methods* **2010**, *2* (11), 1792–1797.

(49) Hunter, E. P.; Lias, S. G. Evaluated Gas-Phase Basicities and Proton Affinities of Molecules: An Update. *J. Phys. Chem. Ref. Data* **1998**, *27* (3), 413–656.

(50) Olivia E. Schroeder; Erica J. Andriole; Krista L. Carver; Kathryn E. Colyer, and; Poutsma*, J. C.; Schroeder, O. E.; Andriole, E. J.; Carver, K. L.; Colyer, K. E.; Poutsma, J. C. Proton Affinity of Lysine Homologues from the Extended Kinetic Method. *J. Phys. Chem. A* **2004**, *108* (2), 326–332.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09. Gaussian, Inc. : Wallingford, CT 2009.

(52) Software, S. PC Model 9. 2006.

(53) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.

(54) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789.

(55) Petersson, G. A.; Al-Laham, M. A. A Complete Basis Set Model Chemistry. II. Open-shell Systems and the Total Energies of the First-row Atoms. *J. Chem. Phys.* **1991**, *94* (9), 6081–6090.

(56) Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100* (41), 16502–16513.

(57) McLean, A. D.; Chandler, G. S. Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, Z=11-18. *J. Chem. Phys.* **1980**, *72* (10), 5639–5648.

(58) Burns, L. A.; Vázquez-Mayagoitia, Á.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: A Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134* (8), 084107.

(59) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241.

(60) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615.

(61) Johnson III, R. D. NIST Computational Chemistry Comparison and Benchmark Database https://cccbdb.nist.gov/credits.asp (accessed Jun 19, 2018).

(62) Buczek, A.; Kupka, T.; Broda, M. A.; Żyła, A. Predicting the Structure and Vibrational Frequencies of Ethylene Using Harmonic and Anharmonic Approaches at the Kohn–Sham Complete Basis Set Limit. *J. Mol. Model.* **2016**, *22* (1), 1–10.

(63) Prazeres, R.; Glotin, F.; Insa, C.; Jaroszynski, D. A.; Ortega, J. M. Two-Colour Operation of a Free-Electron Laser and Applications in the Mid-Infrared. *Eur. Phys. J. D - At. Mol. Opt. Phys.* **1998**, *3* (1), 87–93.

(64) Martens, J.; Berden, G.; Gebhardt, C. R.; Oomens, J. Infrared Ion Spectroscopy in a Modified Quadrupole Ion Trap Mass Spectrometer at the FELIX Free Electron Laser Laboratory. *Review of Scientific Instruments*. AIP Publishing LLC October 17, 2016, p 103108.

(65) Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. Methods and Protocols of Modern Solid Phase Peptide Synthesis. *Mol. Biotechnol.* **2006**, *33* (3), 239–254.

International Journal of Mass Spectrometry 429 (2018) 158–173

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Full Length Article

Infrared multiple photon dissociation spectroscopy of cationized canavanine: Side-chain substitution influences gas-phase zwitterion formation

Zachary M. Smith [a], Vincent Steinmetz [b], Jonathan Martens [c], Jos Oomens [c,d], John C. Poutsma [a,*]

[a] Department of Chemistry, The College of William and Mary, Williamsburg, VA 23187-8795, United States
[b] Laboratoire de Chimie Physique, CNRS UMR 8000, Université Paris Sud, Université Paris Saclay, CNRS, Orsay, France
[c] Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Nijmegen, The Netherlands
[d] Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 15 March 2017
Received in revised form 5 July 2017
Accepted 19 August 2017
Available online 4 September 2017

Dedicated to Professor Terry McMahon on the occasion of his 70th Birthday and in recognition of his outstanding contributions to gas-phase ion chemistry, thermochemistry, and spectroscopy.

Keywords:
Amino acids
IRMPD
Infrared spectroscopy
Gas-phase zwitterions
DFT calculations

ABSTRACT

Infrared multiple photon dissociation spectroscopy was performed on protonated and cationized canavanine (Cav), a non-protein amino acid oxy-analog of arginine. Infrared spectra in the XH stretching region (3000–4000 cm^{-1}) were obtained at the Centre Laser Infrarouge d'Orsay (CLIO) facility. Comparison of the experimental infrared spectra with scaled harmonic frequencies at the B3LYP/6-31+G(d,p) level of theory indicates that canavanine is in a canonical neutral form in $CavH^+$, $CavLi^+$, and $CavNa^+$; therefore, these cations are charge-solvated structures. The infrared spectrum of $CavK^+$ is consistent with a mixture of Cav in canonical and zwitterionic forms leading to both charge-solvated and salt-bridged cationic structures. The Cav moiety in $CavCs^+$ is shown to be zwitterionic, forming a salt-bridged structure for the cation. Infrared spectra in the fingerprint region (1000–2000 cm^{-1}) obtained at the FELIX Laboratory in Nijmegen, Netherlands support these assignments. These results show that a single oxygen atom substitution in the side chain reduces the stability of the zwitterion compared to that of the protein amino acid arginine (Arg), which has been shown previously to adopt a zwitterionic structure in $ArgNa^+$ and $ArgK^+$. This difference can be explained in part due to the decreased basicity of Cav (PA = 1001 kJ/mol) as compared to arginine (PA = 1051 kJ/mol), but not entirely, as lysine, which has nearly the same proton affinity as Cav, (~993 kJ/mol) forms only canonical structures with Na^+, K^+, and Cs^+. A major difference between the zwitterionic forms of $ArgM^+$ and $CavM^+$ is that the protonation site is on the side chain for Arg and on the N-terminus for Cav. This results in systematically weaker salt bridges in the Cav zwitterions. In addition, the presence of another hydrogen-bonding acceptor atom in the side chain contributes to the stability of the canonical structures for the smaller alkali cations.

1. Introduction

As the building blocks of proteins and peptides, amino acids are fundamentally important biological species. In solution at physiological pH, amino acids occur naturally as zwitterions, with a protonated amino terminus and a deprotonated carboxylic acid group [1]. In contrast, the large energy cost to deprotonate amino acids in the gas phase (>1330 kJ/mol) [2,3] is not fully compensated for by the energy savings from protonating the amino terminal group (<1050 kJ/mol) [4,5], leading to canonical ($COOH$, NH_2) structures for gas-phase amino acids. For the simplest amino acid, glycine, the zwitterionic form lies nearly 90 kJ/mol above the lowest energy canonical structure [6]. The stability of the zwitterionic form of isolated gas-phase amino acids is affected by the acidity of the COOH group and the basicity of the amino terminus/side chain. Arginine (Arg) is the most basic of the twenty protein amino acids (PAA) with a PA of 1051 kJ/mol [7,8]. The increased basicity and the availability of multiple hydrogen- bonding sites in arginine helps to stabilize the zwitterionic form, and the difference in energy between the zwitterion and the canonical forms drops to around

* Corresponding author.
E-mail address: jcpout@wm.edu (J.C. Poutsma).

https://doi.org/10.1016/j.ijms.2017.08.009
1387-3806/

15 kJ/mol [9]. Neutral arginine has been shown experimentally to be canonical through infrared cavity ring-down experiments [10].

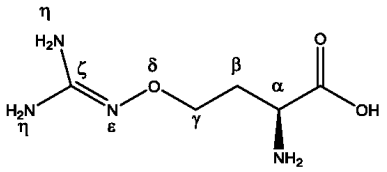
Small perturbations can affect the relative stability of the zwitterionic form of neutral amino acids relative to the canonical form. Addition of water molecules must lead to the zwitterionic forms of the amino acids becoming more stable, and many computational and experimental studies have been performed in order to determine the exact number of water molecules needed to cause the shift in stability [11–22]. A recent study by Perez de Tudela and Marx used spin- component scaled-MP2 to predict that the addition of nine water molecules in a specific bifurcated wire orientation is sufficient to cause the zwitterion form of glycine to be more stable than the canonical structure [22]. This number is between four and five for aliphatic amino acids, drops to around three waters for lysine, and it is predicted that a single water is sufficient to stabilize the zwitterionic form of arginine [23]. Bush et al. [24] showed that a single water molecule is sufficient to stabilize the salt-bridged structure of ArgLi$^+$ relative to the charge-solvated isomer, which is the more-stable species in the gas-phase without the water molecule (see below) [25,26].

Amino acids can form zwitterionic species when complexed to other molecules. The neutral dimer of arginine is predicted to contain one of the neutral arginine molecules in its zwitterionic state [27], and in the proton-bound dimer ion of Arg, hydrogen-deuterium exchange (HDX) [28] and Blackbody Infrared Dissociation (BIRD) [29] spectroscopy have been used to show that one of the neutrals is zwitterionic. In contrast, work by Wu et al. showed that the proton-bound dimer of lysine (less basic by ~50 kJ/mol) is a charge-solvated structure [30].

The richest area of gas-phase amino acid zwitterion research centers on the ability of alkali and other metals to stabilize the zwitterionic form of amino acids through salt-bridged structures. Early ion mobility work from Bowers [31], a BIRD study by Jockush et al., [32] and a kinetic method study from Cerda and Wesdemiotis [33] established that coordination of arginine to increasingly larger alkali metals resulted in binding energies that were more consistent with arginine in a zwitterionic form than in a canonical structure. Early IRMPD studies from the Williams group established that arginine forms a salt-bridged structure when complexed with Na$^+$, K$^+$, and Cs$^+$, but a charge-solvated structure with H$^+$ and Li$^+$ [25,26].

IRMPD spectroscopy has become an invaluable tool for determining the structure of trapped gas-phase ions [34–37]. Tunable infrared radiation from free-electron lasers [38–40] and more recently from bench-top OPO/OPA lasers [25,41–45] has allowed vibrational spectroscopy to be applied to a variety of gas-phase cations and anions. Many recent studies of the structure of ionized amino acids and peptides using IRMPD have been carried out including investigations of their zwitterion/canonical forms [24–26,46–73], low-energy conformations of protonated amino acid clusters [30,51,74–77], and hydration of ionized amino acids [41,42,78–84].

Of particular relevance to this work are recent studies on the structure of protonated and alkali-metallated arginine [25,26] and on the structure of metallated lysine [52,61]. In these studies, IRMPD spectroscopy was used to determine that 1) ArgLi$^+$ is a charge-solvated structure with a canonical arginine moiety, 2) ArgNa$^+$, ArgK$^+$, ArgRb$^+$, and ArCs$^+$ are salt-bridged structures containing a zwitterionic arginine moiety, 3) LysM$^+$ forms exclusively charge-solvated structures, 4) *N*-methylation of lysine at either the amino terminus (α-N) or the side chain (ε-N) stabilizes the zwitterionic form of the lysine moiety such that MeLysNa$^+$ and MeLysK$^+$ are salt-bridged structures, and 5) dimethylation of the side chain results in exclusively salt-bridged structures for M = Li$^+$, Na$^+$, and K$^+$. The PA of Lys has been measured to be between 988 and 1007 kJ/mol [85–88], with recent high-level calculations centering in on a value of 993 kJ/mol [4,5,61]. Bush et al. calculated proton affinities for ε-NMeLys, α-NMeLys, and Me_2Lys to be 997, 1008, and 1014 kJ/mol, respectively. The PA for arginine has recently been re-determined to be much higher than the lysine analogs (1051 kJ/mol) [8]. These studies reveal that the proton affinity of the isolated amino acids in and of itself is a poor predictor for whether they will adopt canonical or zwitterionic structures when complexed to alkali metals as the only salt-bridged cation containing lithium is the Me_2LysLi$^+$. Instead, one must consider the basicity in concert with the hydrogen-bonding ability of the side chain of the amino acid, the identity of the ionized groups, and the different metal binding motifs available to the amino acid. Given these results, we were inspired to investigate the structures of a series of metallated cations of canavanine, an oxy-analog of arginine with an oxyguanidino group as its side chain.



Scheme 1. Naming scheme for atomic sites in canavanine.

We have been interested in the gas-phase chemistry of the so-called "non-protein" amino acids (NPAA), which are naturally occurring species that are not used for protein/peptide synthesis [89,90]. These compounds are ubiquitous in nature and are often found as secondary products of plant metabolism. Many NPAAs are structurally similar to one or more of the 20 protein amino acids (PAA) and can compete with them in a variety of biochemical processes including mis-incorporation into proteins [91–106]. In addition to their biological significance, these species are excellent models for examining the subtle interplay between structure and energetics in amino acids. We have measured gas-phase proton affinities [107] and acidities [108] for proline-analogs using the extended kinetic method and recently showed that the six-membered ring analog, pipecolic acid, leads to a selective fragmentation mechanism when inserted into peptides [109]. We have also determined proton affinities and gas-phase acidities of homologs of lysine, serine, and cysteine [87,110].

In 2006, we measured the proton affinity of canavanine (Cav, Scheme 1), an oxy-analog of the PAA arginine, that results from oxygen atom substitution for the δ-CH_2 group [111]. Using the extended kinetic method in a quadrupole ion trap instrument, we determined a PA for Cav of 1001 ± 9 kJ/mol. Thus, the single atom substitution in the side chain results in a nearly 50 kJ/mol reduction in proton affinity. This effect has been observed in solution, and the pKa of the oxyguanidino side-chain has been measured to be ~7 units [93,112]. Because of the similarity in structure between Cav and Arg, most organisms' t-RNA synthetase molecules cannot differentiate between them, which allows for facile mis-incorporation of Cav into proteins [102,106,113,114]. Cav is a potent natural insecticide [102,106,115], has been investigated for use as an anti-cancer drug [113,116], and has been shown to increase the potency of other anti-cancer therapies [117–119]. The decrease in basicity of the side chain should result in less stable zwitterionic structures for Cav relative to Arg, and the addition of another hydrogen-bonding acceptor atom in the side chain should allow for different hydrogen-bonding schemes to be present in Cav. We present here a combined experimental/computational study of the structure of CavH$^+$, CavLi$^+$, CavNa$^+$, CavK$^+$, and CavCs$^+$ and show that the behavior of the CavM$^+$ ions is intermediate between that of

ArgM$^+$ (salt- bridged structure for Na^+, K^+, and Cs^+) and that of LysM$^+$ (charge-solvated structure for all cations).

2. Experimental section

2.1. CLIO

IRMPD spectra for several species were obtained using the CLIO free- electron laser (FEL) in Orsay, France. An infrared spectrum for CavH$^+$ in the fingerprint region (~1000–2000 cm^{-1}) was obtained using the FEL in a modified Bruker Esquire quadrupole ion trap instrument [120,121]. The laser light from the FEL comes in 8 μs-long macropulses at a repetition rate of 25 Hz. The average laser power was about 500 mW, which corresponds to a micropulse energy of 40 μJ, and a macropulse energy of 20 mJ. The wavelength range for the free electron laser was tuned between *ca.* 1000–2000 cm^{-1}. The laser bandwidth is about 0.3–0.5% of the center wavelength. The wavelength scale was calibrated by passing a portion of the beam through a polystyrene film and recording the infrared spectrum in the region between 1000 and 2000 cm^{-1} in real time. The experimental polystyrene spectrum is fit to the known spectrum using a linear fit, which is then applied to the wavelength data for the analyte. Spectra are also linearly corrected for variations in the laser power in different regions of the spectrum. Infrared spectra for CavH$^+$, CavLi$^+$, CavNa$^+$, CavK$^+$, and CavCs$^+$ in the X—H stretching region (*ca.* 3200–3600 cm^{-1}) were obtained using a Nd:YAG pumped OPO/OPA system coupled to a modified Bruker ion cyclotron resonance instrument that has been described in detail previously [122,123]. The OPO/OPA laser was pumped by a 25 Hz Nd:YAG laser operating at 550 mJ/pulse with a 4–6 ns pulse duration at 1064 nm. Typical output energy is ~12 mJ/pulse with a spectral width of ~5 cm^{-1}.

Analyte spectra are obtained as "action" spectra in which the infrared light from the FEL or OPO is allowed to interact with trapped ions. CavH$^+$ was formed by directly infusing a dilute solution (~10 μM) of Cav in acidified (~1% formic acid) 50:50 H_2O:CH_3OH into the electrospray ionization source. Alkali-metallated Cav ions were similarly formed from electrospray ionization of an approximately 3:1 mixture of MCl (X = Li, Na, K, Cs) and Cav in the same solvent system at a total concentration of ~10 μM for Cav and ~30 μM for X$^+$. Ion focusing and mass spectrometer tuning conditions are optimized for the intensity of CavH$^+$ or CavM$^+$ precursor ions. Analyte ions are mass selected and are then allowed to interact with the tunable infrared light. If the laser light is resonant with one of the vibrations in the ion, multiple photons are absorbed, and the ion can fragment. A spectrum is generated by plotting the IR yield (the sum of all product ions normalized to the total ion intensity, Σ(products)/Σ(products + parent)).

The fingerprint region spectrum for CavH$^+$ was obtained on the Bruker Esquire ion trap instrument. The parent ion (*m/z* = 177 amu) fragments to give peaks at *m/z* = 160 (loss of NH_3), 159 (loss of H_2O), 118 (loss of guanidine NH═C(NH_2)$_2$), and 76, which is probably protonated oxyguanidine (HON(H)═C(NH_2)$_2$). X—H stretching region spectra for CavH$^+$, CavLi$^+$, CavNa$^+$, CavK$^+$, and CavCs$^+$ were obtained on the ion cyclotron resonance instrument. For CavH$^+$, CavLi$^+$, and CavCs$^+$, fragmentation was enhanced by non-resonant absorption of IR light from a CO_2 laser (3 ms duration) followed by 12 pulses from the OPO laser light. For CavNa$^+$ and CavK$^+$ the number of pulses was doubled.

Under these conditions, CavH$^+$ gives fragments at *m/z* 159 and 160 amu, corresponding to loss of H_2O and NH_3, respectively. For CavLi$^+$, the parent ion (*m/z* 183) dissociates only through loss of water to give a fragment ion of *m/z* = 165 amu. CavNa$^+$ (*m/z* 199) fragments to give peaks at *m/z* 181 and 182 amu, corresponding to loss of H_2O and NH_3. The parent ion for CavK$^+$ at *m/z* 215 amu fragments only through loss of ammonia (*m/z* 198) under these conditions. Finally, CavCs$^+$ (*m/z* 309) fragments to give peaks at *m/z* 292 (loss of NH_3), *m/z* 133 (Cs$^+$), and *m/z* 103 (loss of the oxyguanidino group $^•$ON═C(NH_2)$_2$).

2.2. FELIX

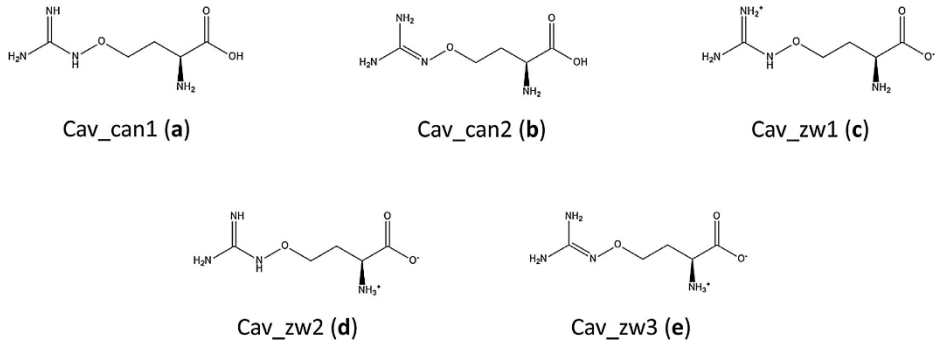
Similar procedures were used to obtain fingerprint-region (~900–2000 cm^{-1}) spectra for CavNa$^+$, CavK$^+$, and CavCs$^+$ at the FELIX Laboratory using a modified Bruker amaZon ion trap mass spectrometer [124]. Unfortunately, we were unable to obtain a fingerprint spectrum for CavLi$^+$ during our allotted beam time at FELIX. For these studies, much less concentrated (0.1 μM) solutions were used and the ratio of alkali metal salt to canavanine was closer to 1:1. Ion source conditions were optimized for CavM$^+$ intensities. Precursor ions are isolated in the ion trap and allowed to interact with tunable light from the free electron laser. The laser light arrives in 5 μs macropulses with approximate energy of 30–40 mJ/pulse. The laser bandwidth is approximately 0.4% of the center frequency. The wavelength is calibrated online using a grating spectrometer, and a linear correction for frequency-dependent variation in laser power is applied.

In the fingerprint region, multiple-photon absorption on CavNa$^+$ (*m/z* 199) produces fragments with *m/z* = 182 (loss of NH_3), *m/z* 181 (loss of H_2O), *m/z* 139 (loss of 60: H_2O and NH═C═NH) [125] and *m/z* 98 (which is probably sodiated oxyguanidine, Na^+–HO—N═C(NH_2)$_2$). CavK$^+$ (*m/z* = 215) fragments to lose NH_3 (*m/z* = 198) and CO_2 (*m/z* 171). Interestingly, the spectra for these individual ions losses have very different features suggesting a mixture of parent ion structures (see below). For CavCs$^+$, the only fragment ion has a mass of 249, which corresponds to a loss of 60: H_2O and NH═C═NH [125].

2.3. Theoretical procedures

Calculations for Cav, CavH$^+$, CavLi$^+$, CavNa$^+$, CavK$^+$, and CavCs$^+$ were performed using the Gaussian09 suite of programs [126]. The GMMX conformer searching routine in PCModel [127] is used to find conformations within 40 kJ/mol of the minimum-energy structure for Cav in its two canonical tautomer-classes and its three zwitterionic isomers (see Scheme 2 below). These structures are used as starting points for a series of molecular orbital and density functional theory calculations at progressively higher levels of theory. Ultimately, geometries, zero-point energies, and thermal corrections to enthalpy and free energy for all species except CavCs$^+$ were calculated at the B3LYP/6-31+G(d) level [128,129]. For CavCs$^+$, we used the LANL2DZ [130–133] basis set for geometry optimization, zero-point energies, thermal corrections to enthalpy and free energy, and harmonic vibrational frequencies. At the request of a reviewer, we re-optimized the 15 lowest energy CavCs$^+$ conformers and 13 additional CavCs$^+$ conformers from the LANL2DZ study using the DEF2 -TZVPP basis set. Similarities and differences between the energetics and spectra obtained from the two basis sets are discussed in the CavCs$^+$ section (see below).

Zero-point energy (ZPE) and thermal corrections for all species were obtained from un-scaled harmonic vibrational frequencies. As the recommended scaling factors for B3LYP frequencies for ZPE and thermal corrections are 0.98 and 0.99 respectively [134], we chose to use un-scaled frequencies for deriving ZPE and thermal corrections. To determine the relative stabilities of the different conformers of the various isomers, total electronic energies were obtained for all species except CavCs$^+$ using single-point energy calculations at the B3LYP/6-311++G(d,p) level and converted to 298 K enthalpies and free energies using corrections from the B3LYP/6-31+G(d) frequencies.



Scheme 2. Tautomeric forms of canonical and zwitterionic neutral canavanine.

In order to make comparisons to previously published work and to help guide our interpretation of results, similar calculations were carried out for Arg, $ArgH^+$, $ArgLi^+$, $ArgNa^+$, $ArgK^+$, $ArgCs^+$, Lys, $LysH^+$, $LysLi^+$, $LysNa^+$, $LysK^+$, and $LysCs^+$ at the same levels of theory as those for the ions containing Cav. Tabulated values for these calculations are not presented here, though the results of selected calculations are referred to throughout the manuscript. Full computational results for the arginine- and lysine-related calculations are available upon request.

Geometries for selected low-energy conformers were re-optimized at the B3LYP/6-31+G(d,p) level of theory in order to obtain calculated vibrational spectra for comparison to the IRMPD spectra. Scaling factors for the calculated harmonic frequencies in the fingerprint region were 0.975[135] for $CavH^+$, $CavNa^+$, and $CavK^+$ and 0.98 for $CavCs^+$. In the X—H region of the spectra, the scaling factors were 0.955 ($CavH^+$, $CavLi^+$, and $CavNa^+$), 0.960 ($CavK^+$), and 0.965 for ($CavCs^+$) [135]. Stick spectra were broadened by applying a 20 cm^{-1} baseline-width Gaussian function.

3. Materials

Canavanine, LiCl, NaCl, KCl, and CsCl were purchased from Sigma Aldrich (St. Louis, MO) and were used as provided with no further purification.

4. Results and discussion

4.1. Canavanine and protonated canavanine

The oxyguanidino group in Cav has five tautomers of two general classes: can1 and can2. Four tautomers have the cav_can1 (**a**, Scheme 2) connectivity, which has a hydrogen atom on the ε-nitrogen (adjacent to the oxygen) and a formal double bond between the ζ-carbon and one of the η- nitrogen atoms (see Scheme 1 for lettering of Cav sites). These four tautomers comprise two sets of two *cis/trans* isomers that differ in terms of which η-nitrogen has the double bond. All four tautomers were considered in the GMMX search as we allowed rotation of both the O—N single bond and the ζC=ηN double bond. The second tautomer class, cav_can2 (**b**) comprises only one tautomer and has no hydrogen on the ε-nitrogen, a formal double bond between the ε-nitrogen and ζ-carbon, and two hydrogens on each of the η-nitrogen atoms (Scheme 2). The two tautomer classes lead to three different zwitterionic neutrals: cav_zw1 (**c**), which has a neutral amino group, a deprotonated carboxylic acid group, and a protonated side chain, cav_zw2 (**d**), which has a protonated amino group, a deprotonated acid group, and a can1 tautomer neutral side chain, and cav_zw3 (**e**), that has a protonated amino group, deprotonated acid group, and the can2 tautomer neutral side chain. The three zwitterions for neutral canavanine are also shown in Scheme 2. For ease of notation, we will be using **a**–**e** to describe the neutral form of the canavanine moiety in the various complexes according to Scheme 2. Due to the fact that the some of the tautomers isomerized during optimization, the following numbering scheme was developed for the Figures and Tables: all optimized structures for a given molecule were sorted by 298 K free energy and all conformers are assigned numbers according to this ranking. The final conformer type (**a**–**e**) is appended to the structure name in parentheses. Thus the global minimum Cav structure is denoted cav_001 (**b**) indicating that it is canonical and has its side chain in the can2 tautomer.

Fig. 1 shows the lowest free-energy conformer for each of the five neutral isomers for Cav calculated at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory. Conformers of Cav with the side chain in the can2 tautomer have the lowest free energy at 298 K, followed by those with the side chain in the can1 tautomer class. The lowest can1 isomer is 35.1 kJ/mol higher in free energy than the global minimum structure. The lowest free-energy zwitterion is of structure zw3 (**e**) and lies 55.1 kJ/mol above the global minimum canonical structure. The lowest lying zwitterion of form zw1 (**c**) was found to be 57.0 kJ/mol above the lowest energy canonical structure. All but one of the 26 starting structures for zw2 (**d**) generated by our searching routine isomerized to can1 (**a**) isomers through proton transfer from the protonated amino group to the carboxylate oxygen. The one remaining zw2 (**d**) isomer lies 78.9 kJ/mol above the global minimum structure. In all, our conformational search identified a total of 166 canonical conformers within 86.3 kJ/mol of the global minimum structure and 10 zwitterionic conformers ranging from between 55.1 and 104.8 kJ/mol above the global minimum. Table 1 shows the relative free-energy values for the lowest lying conformer of each type for Cav (**a**–**e**), $CavH^+$ (**i**–**iv**), and $CavM^+$ (M = Li, Na, K, and Cs; **a**–**e**). Fig. S1 of Supporting Information shows the structures of six additional low-lying conformers of Cav (**a**–**e**). Table S1 of Supporting Information gives a list of the relative free energies at 298 K for all conformers of neutral canavanine found in our search.

As Cav is an oxy-analog of the protein amino acid arginine, it is useful to compare the structural features of Cav with those of Arg. There have been numerous computational studies of neutral arginine, many of which were supporting experimental studies of its thermochemistry or gas-phase structure [3–5,8,9,25,26,29,111,136]. The relative stability of the can1 and can2 tautomers depends on the basis set, with the highest levels of

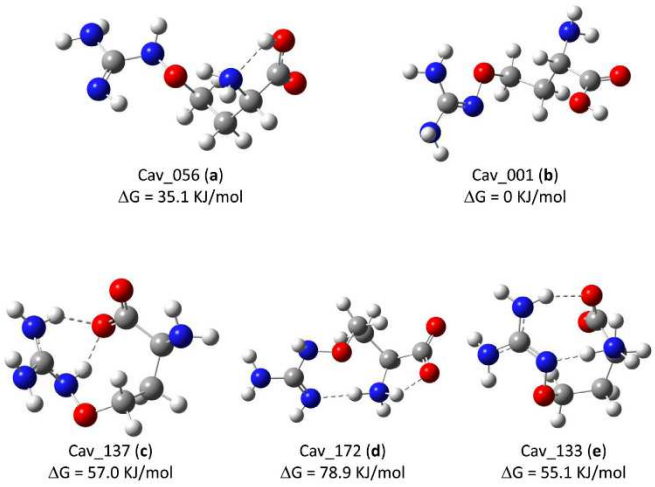


Fig. 1. Lowest free-energy conformers for Cav (**a–e**) calculated at the B3LYP/6-31++G (d,p)//B3LYP/6-31+G(d) level of theory.

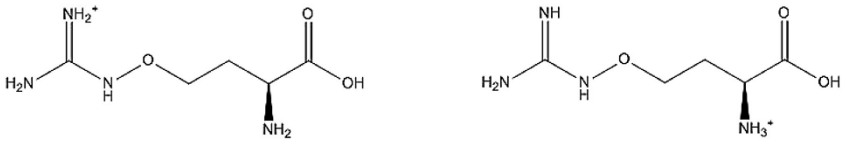
Table 1
Calculated 298 K enthalpies, 298 K free energies (hartrees) and relative free energies (kJ/mol) for Cav, CavH$^+$, and CavM$^+$ (M = Li, Na, K, Cs).

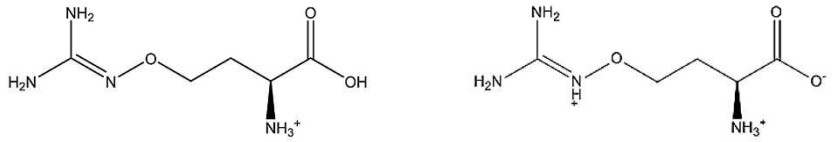
Molecule	isomer	H^0_{298}	G^0_{298}	ΔG
		Cav		
Cav_001 (**b**)	**can2**	−642.398373	−642.455045	0.0
Cav_056 (**a**)	**can1**	−642.386825	−642.441687	35.1
Cav_133 (**e**)	**zw3**	−642.383534	−642.434047	55.1
Cav_137 (**c**)	**zw1**	−642.381223	−642.433319	57.0
Cav_172 (**d**)	**zw2**	−642.372391	−642.424967	79.0
		CavH$^+$		
CavH_001 (**i-1**)	**i-1**	−642.779338	−642.835585	0.0
CavH_005 (**i-2**)	**i-2**	−642.775826	−642.832079	9.2
CavH_006 (**i-3**)	**i-3**	−642.777949	−642.831749	10.1
CavH_011 (**i-4**)	**i-4**	−642.777110	−642.830578	13.1
CavH_045 (**iii**)	**iii**	−642.770779	−642.825354	26.9
CavH_059 (**i-5**)	**i-5**	−642.766506	−642.820039	40.8
CavH_064 (**iv**)	**iv**	−642.757876	−642.811234	63.9
		CavLi$^+$		
CavLi_001 (**a**)	**a**	−649.818625	−649.872269	0.0
CavLi_002 (**b**)	**b**	−649.815889	−649.870834	3.8
CavLi_044 (**c**)	**c**	−649.800023	−649.856806	40.6
CavLi_046 (**e**)	**e**	−649.800487	−649.856680	40.9
CavLi_068 (**d**)	**d**	−649.796955	−649.851682	54.1
		CavNa$^+$		
CavNa_001 (**b**)	**b**	−804.580085	−804.637621	0.0
CavNa_003 (**a**)	**a**	−804.578427	−804.635014	6.8
CavNa_006 (**e**)	**e**	−804.574721	−804.631782	15.3
CavNa_014 (**c**)	**c**	−804.571501	−804.628213	24.7
CavNa_023 (**d**)	**d**	−804.569845	−804.626525	29.1
		CavK$^+$		
CavK_001 (**e**)	**e**	−1242.226037	−1242.285040	0.0
CavK_004 (**b**)	**b**	−1242.225111	−1242.284518	1.4
CavK_006 (**c**)	**c**	−1242.222194	−1242.280507	11.9
CavK_008 (**d**)	**d**	−1242.221063	−1242.279349	14.9
CavK_026 (**a**)	**a**	−1242.216342	−1242.275098	26.1
		CavCs$^+$		
CavCs_001 (**e**)	**e**	−662.526112	−662.588024	0.0
CavCs_003 (**b**)	**b**	−662.522891	−662.584703	8.7
CavCs_005 (**c**)	**c**	−662.523185	−662.583196	12.7
CavCs_007 (**d**)	**d**	−662.521755	−662.582150	15.4
CavCs_025 (**a**)	**a**	−662.510320	−662.573641	37.8

the theory predicting that the can2 tautomer is the lower energy structure [5,9,137]. At the B3LYP/6- 311++G(d,p)//B3LYP/6-31+(d) level of theory the can2 tautomer is 7.8 kJ/mol more stable than the can1 tautomer. Despite the large proton affinity for arginine, the canonical structure for neutral arginine is predicted to be *ca.* 15 kJ/mol more stable than the zwitterionic form [9], which is a much smaller difference than that predicted here for canavanine.

Calculations were also performed for protonated canavanine (CavH$^+$). Protonating canavanine leads to three different canonical structures: protonation on the side chain (**i**), protonation on the N-terminal amino group with a can1 neutral side chain (**ii**), and protonation on the N-terminal amino group with a can2 neutral side chain (**iii**). In addition, a single zwitterionic cation can be formed by deprotonating the carboxylic acid group and protonating both the side chain and the N-terminal amino group (**iv**). Structures for these isomers are shown in Scheme 3. The lowest-energy isomers for CavH$^+$ are canonical and are of type (**i**). Two nearly iso-energetic ($\Delta\Delta G$ = 0.02 kJ/mol) structures were found both possessing strong hydrogen bonds between the hydrogen on the ε-nitrogen atom and the amino terminus (this hydrogen-bonding scheme is denoted **i-1**, Fig. 2). Though they share the same hydrogen-bonding motif, they differ in the relative conformation of the seven-membered "ring" formed from the intramolecular hydrogen bond. Interestingly, though these two isomers are nearly equal in free energy, they have a difference in H_{298} at this level of more than 7 kJ/mol. Fig. 2 shows the global minimum structure along with other low-lying canonical and zwitterionic isomers for CavH$^+$. Table S2 in Supporting Information lists the relative free energies for all isomers of CavH$^+$ found in the conformational search. The structures and relative free energies of eight additional conformers of CavH$^+$ **(i–iv)** are shown in Fig. S2 of Supporting Information.

In addition to the **i-1** conformers described above, we located several additional **i**-type conformers with different internal hydrogen- bonding motifs: isomer (**i-2**) with a hydrogen bond between the hydrogen on the *ε-nitrogen* and the carbonyl oxygen, and isomers (**i-3**), (**i- 4**), and (**i-5**) with hydrogen bonds between the hydrogen atom on the *η-nitrogen* and either the amino terminus, carbonyl oxygen, or hydroxyl oxygen, respectively. The lowest energy conformers of these are shown in Figs.





Scheme 3. Tautomeric forms of canonical and zwitterionic protonated canavanine.

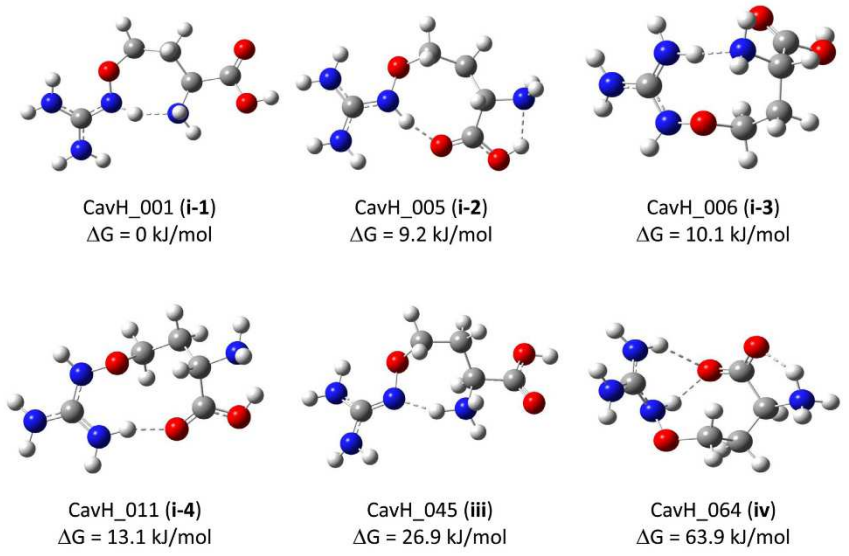


Fig. 2. Lowest free energy conformers for $CavH^+$ (**i-1 – i-4**, **iii**, and **iv**) calculated at the B3LYP/6-31++G(d,p)//B3LYP/6-31++G(d) level of theory.

2 and S2. Several type- **iii** conformers were found, the lowest of which lies 26.9 kJ/mol above the global minimum structure. All type- **ii** conformers isomerized during the geometry optimization through proton transfer from the protonated amino terminus to the η-nitrogen of the side chain. Several high-lying zwitterionic conformers (**iv**) were located, the lowest of which lies 64 kJ/mol above the global minimum structure and is shown in Fig. 2.

Our current conformational searching routine makes use of larger energy windows in the initial GMMX search than we used in some of our earlier work, including our study of the proton affinity of canavanine [111]. Consequently, many more initial conformations are included in our searching routine, which lends increased confidence that our search is finding all important minima on the potential energy surface. When we published our study on the proton affinity of canavanine in 2006, we a) used a smaller energy window in the GMMX search (12 kJ/mol instead of 40 kJ/mol), b) did not account for the possibility of zwitterions, and c) used only the lowest *enthalpy* conformer of the cation and neutral to determine our proton affinity prediction [111]. It has since become standard practice in our lab to present proton affinities based on Boltzmann-weighted enthalpies for the neutral and protonated amino acids under study [110,138]. We use G_{298} to determine the Boltzmann population for each low-energy conformer to determine the weighting factor for the H_{298} values. In this case, we were pleased to find that the lowest enthalpy neutral conformer that we based our proton affinity value on in 2006 is the lowest free-energy conformer from our current more robust search. Interestingly, we did locate several additional conformers that were between 1 and 4 kJ/mol lower in enthalpy, but these species have higher free energies than the global minimum structure. In addition, the lowest enthalpy conformer for $CavH^+$ from our previous study is still the lowest enthalpy structure in the current study, though it is the less stable of the two lowest free-energy isomers (by 0.01 kJ/mol). Using the relative ΔG_{298} values for Cav and $CavH^+$ to generate 298 K Boltzmann weighting factors leads to a raw PA of 1007.8 kJ/mol for Cav. Using ethylenediamine (PA = 951.4 kJ/mol) [139] as an isodesmic reference gives a slightly larger value of 1008.2 kJ/mol for the PA of Cav. This value is in better agreement with our experimental determination of 1001 ± 9 kJ/mol for the PA of Cav than the non-Boltzmann-weighted prediction of 1015 kJ/mol from our previous work [111] and in near quantitative agreement with a recent calculated Boltzmann-weighted proton affinity for Cav by Rozman at the G3MP2B3 level of theory of 1009.5 kJ/mol [140].

Our calculations predict that, as with all other amino acids, the gas-phase structure for $CavH^+$ should be canonical rather than zwitterionic. To test this experimentally, we obtained an IRMPD action spectrum of $CavH^+$ in the X—H region of the spectrum (3200–3650 cm^{-1}) using the OPO/ICR setup at CLIO. The infrared spectrum is shown in the bottom trace of Fig. 3. Fig. 4 shows the experimental spectrum in the X—H region along with calculated spectra for various low-lying conformers of $CavH^+$. There is a strong peak in the experimental spectrum near 3570 cm^{-1} with a slightly less intense shoulder at 3550 cm^{-1} and a smaller peak around 3460 cm^{-1}. An examination of Fig. 4 shows that the experimental spectrum matches very well with the calculated harmonic spectra for the type **i-1** conformer, with an internal hydrogen bond between a hydrogen atom on the ε-nitrogen of the protonated side chain and the N-terminal amino nitrogen (Fig. 4b). The calculated spectra for canonical isomers (**i-2**), (**i-3**), (**i-4**), and (**iii**) shown in Fig. 4c–f and the zwitterionic form of $CavH^+$ (**iv**) shown in Fig. 4g do not match the experimental spectrum as well as that of **i-1**.

An examination of the normal modes from the lowest-energy conformer (Fig. 4b) indicates that the wide peak and shoulder near 3550–3570 cm^{-1} actually arises from three different modes, two different antisymmetric NH_2-stretching modes for the two η-NH_2 groups, and the free O—H stretch. The smaller peak near 3460 cm^{-1} also arises from two normal modes, the two symmetric NH_2-stretching modes for the two η-NH_2 groups. Fig. 4b shows the calculated spectrum for conformer **i-2** that has an internal hydrogen bond between the hydrogen atom on the ε-nitrogen atom on the protonated side chain and the carbonyl oxygen (see Fig. 2 for structure). For this conformer, the OH group forms a hydrogen bond

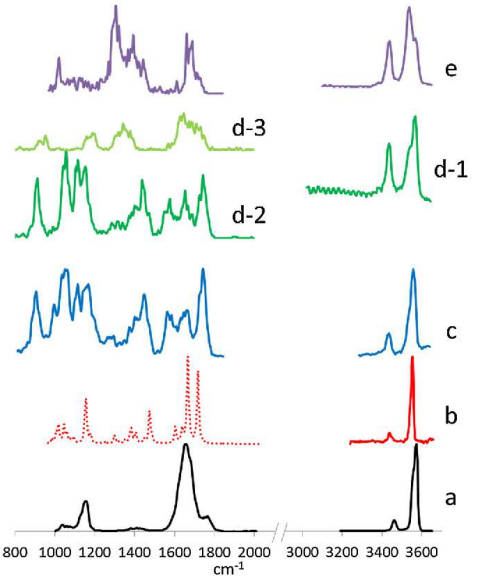


Fig. 3. IRMPD spectra for **a**) $CavH^+$ in the fingerprint and X—H regions, **b**) $CavLi^+$ in the fingerprint (calculated) and X—H regions, **c**) $CavNa^+$ in the fingerprint and X—H regions, **d-1**) $CavK^+$ in the X—H region, **d-2**) $CavK^+$ in the fingerprint region: m/z = 198 (loss of NH_3, **d-3**) $CavK^+$ in the fingerprint region: m/z = 171 (loss of CO_2) and **e**) $CavCs^+$ in the fingerprint and X—H regions. X—H spectra were obtained at CLIO, fingerprint spectra obtained at FELIX except for $CavH^+$, which was obtained at CLIO.

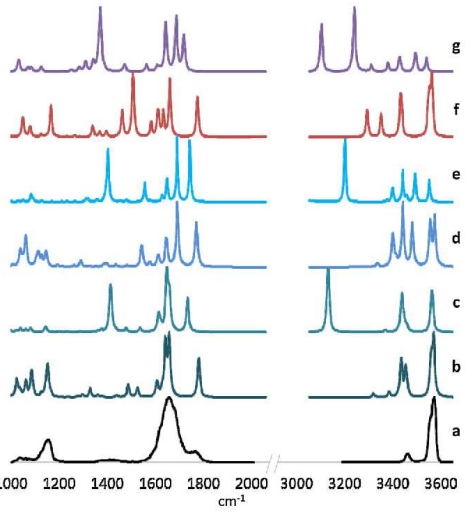


Fig. 4. **a**) Experimental fingerprint and X—H region IRMPD spectrum for $CavH^+$, and calculated infrared spectra for **b**) CavH_001 (**i-1**), **c**) CavH_005 (**i-2**), **d**) CavH_006 (**i-3**), **e**) CavH_011 (**i-4**), **f**) CavH_045 (**iii**), **g**) CavH_064 (**iv**).

with the amino terminal nitrogen atom, which red shifts the O—H stretching frequency down to *ca.* 3130 cm^{-1}. The peak at 3500 cm^{-1} for this conformer narrows considerably and the relative intensity of the two N—H stretching peaks (3550 and 3450 cm^{-1}) becomes more equal. The calculated spectrum for this conformer shows that the two N—H stretching peaks of the oxyguanidine group should be of nearly equal intensity, indicating that the large peak at 3550–3570 cm^{-1} in the experimental spectrum arises mainly from the free O—H stretching mode of the canonical cation.

The experimental infrared spectrum of $CavH^+$ is virtually identical in this region of the spectrum to that of $ArgH^+$ from Bush et al. [25]. In that study on the IRMPD spectroscopy of protonated and alkali- metallated arginine, they cited the intensity of the free O—H stretching vibration in the spectra of $ArgH^+$ and $ArgLi^+$ as evidence for a canonical structure for $ArgH^+$ and a charge-solvated structure for $ArgLi^+$ [25]. The decreased intensity of this peak (and the subsequent equalization of the intensities of the two peaks at 3540 and 3420 cm^{-1}) in the spectra of $ArgNa^+$ and $ArgK^+$ was presented as evidence of a shift to salt-bridged structures containing a zwitterionic arginine [25].

As a further confirmation of the structure of $CavH^+$, we obtained its IRMPD action spectrum in the fingerprint region (1000–2000 cm^{-1}) using the FEL/ion trap setup at CLIO. This spectrum is also shown in the bottom trace of Fig. 3. A comparison of the experimental spectrum with calculated spectra of the low-lying $CavH^+$ isomers in the fingerprint region is also given in Fig. 4. The experimental spectrum shows three strong features: a peak near 1770 cm^{-1} corresponding to the carbonyl C=O stretch, a broad peak from 1590 to 1720 cm^{-1} arising from various N—H bending modes of the side chain oxyguanidine group and the amino terminus, and a bending mode near 1150 cm^{-1}. As with the X—H region of the spectrum, the experimental fingerprint region matches well with the **i-1**- type canonical structure. The calculated spectra for the other canonical isomers (**i-2**), (**i-3**), (**i-4**), and (**iii**) do not match the experimental spectrum, nor does the calculated spectrum for the zwitterion (**iv**). The fingerprint spectrum of $CavH^+$ is nearly identical to that of $ArgH^+$ taken by Forbes et al. at FELIX [26]. The presence of a C=O stretch around 1750 cm^{-1} in their spectrum of ArgOMe+H^+ (which cannot form a zwitterion) was taken as evidence of a canonical COOH group. In the zwitterion, the COOH group is replaced by a COO— with a concomitant red shift of the asymmetric C=O stretch. In pioneering work in this area, Kapota et al. showed that this asymmetric C=O peak in sodiated proline shifts to below 1700 cm^{-1} [46]. In $CavH^+$, the calculations predict a smaller shift (~1720 cm^{-1}) as this peak is coupled to several of the N—H bending modes. The presence of the peak near 1780 cm^{-1} in the experimental $CavH^+$ spectrum reinforces the assignment of the canonical isomer.

Given the agreement between the $CavH^+$ and $ArgH^+$ experimental spectra in the X—H and fingerprint regions, the previous assignment of a canonical structure for $ArgH^+$, and the agreement between the experimental spectra for $CavH^+$ and the calculated spectra for low-lying canonical structures in both the X—H and fingerprint regions, we are confident that $CavH^+$ is a canonical cation under our experimental conditions.

4.2. $CavNa^+$ and $CavLi^+$

For calculations on the lithiated, sodiated, potassiated, and cesiated canavanine, the alkali metal was complexed with all five different neutral forms of Cav. The lowest free-energy conformer for each of the five isomers of $CavNa^+$ is shown in Fig. S3, whereas Fig. S4 shows eight additional low- lying conformers for $CavNa^+$ (**a–e**). For $CavNa^+$, the lowest-energy conformer has the canavanine moiety in the can2 tautomer (**b**), leading to a charge-solvated structure for the cation as shown in Fig. S3. This conformer has the sodium ion interacting with the carbonyl oxygen atom (2.3 Å), the amino-terminal nitrogen atom (2.4 Å), and the oxygen atom (2.5 Å) and ε-nitrogen atom (2.3 Å) of the side chain. A nearly isoenergetic conformer (+0.9 kJ/mol) has the same interactions with a slightly different side-chain carbon skeletal conformation (Fig. S4). The lowest energy charge-solvated isomer with canavanine in the can1 tautomer (**a**) lies 6.8 kJ/mol higher in energy and has the sodium ion interacting similarly with the carbonyl oxygen atom (2.3 Å), the amino terminal nitrogen atom (2.4 Å), and the side-chain oxygen atom (2.5 Å), but with the η-nitrogen atom (2.3 Å) of the side chain. The lowest-energy salt-bridged cation has the canavanine moiety in the type zw3 (**e**) and lies only 15.3 kJ/mol above the global minimum charge-solvated structure. Higher energy salt-bridged structures with the canavanine in the zw1 (**c**) and zw2 (**d**) isomers lie 24.7 and 29.1 kJ/mol above the global minimum, respectively. Table S3 gives the relative free energies for all 317 charge-solvated and salt-bridged conformers for $CavNa^+$ found in the conformational search.

In their previous IRMPD study, Bush et al. did extensive modelling of $ArgNa^+$ and concluded that the global minimum structure at the B3LYP/LACVP++** level is a salt-bridged isomer (structure D in their nomenclature), with the lowest lying charge-solvated isomer (their structure A) lying 10 kJ/mol higher in free energy at 410 K [25]. All of their structures show arginine with the can1 tautomer of the side chain, so their salt-bridged structure is most comparable to our $CavNa^+$ zw1 tautomer (**c**), which lies 9.4 kJ/mol above the lowest free-energy zw3 isomer (**e**) for $CavNa^+$. Their lowest free-energy non-zwitterionic structure (A in their nomenclature) is virtually identical to the (**a**)-type isomer for $CavNa^+$, except for the lack of the additional interaction of the sodium ion with the oxygen atom in the side chain. Clearly, oxygen atom substitution in Cav has a large effect on the charge distribution in the side chain, and the decreased basicity of the oxyguanidino group provides less stability to salt-bridged structures. Additionally, the preference for the can2 tautomer and the presence of an additional hydrogen-bond acceptor/alkali metal ion binding site in Cav allows for energetically-favorable interactions to occur in the can2 configuration, stabilizing charge-solvated structures relative to the salt-bridged structures. Ultimately this leads to a destabilization of the zwitterionic isomer for canavanine relative to the canonical isomers (+15.3 kJ/mol for $CavNa^+$ vs. −13.5 kJ/mol for $ArgNa^+$ at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory).

As our computational procedures are slightly different than previous studies with respect to conformational searching routine and basis sets [25,26,52,61], we performed full computational studies on all of the $ArgM^+$ species in order to verify that our methods identified all of the conformers that were located previously and to account for the can2 and zw3 isomers containing the other side-chain tautomer. Table S4 gives the free energy values for the lowest lying isomers of each type for Arg, $ArgH^+$, and $ArgM^+$ (M = Li^+, Na^+, K^+, Cs^+). Our calculations for $ArgNa^+$ at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory agree qualitatively with those of Bush et al. [25]. The difference in 298 K free energy between our lowest-energy salt-bridged isomer at this level (type **c**, their structure D) and our lowest-energy charge-solvated structure is 13.5 kJ/mol. Interestingly, this charge-solvated conformer (type **a**, their structure B) is 5 kJ/mol lower in free energy than the structure A minimum found by Bush. The difference in the relative free energies of the different charge-solvated structures and salt-bridged structures with the Bush et al. study is a function of the different basis sets used for geometry optimizations and single point energies. The fact that our approach correctly orders the zwitterionic/non-zwitterionic isomers for $ArgNa^+$ gives us confidence in our computational prediction that $CavNa^+$ should be canonical in the gas phase.

As mentioned previously, Bush et al. presented IRMPD spectra for $ArgNa^+$ and $ArgLi^+$ and indicated that the strong intensity of the free OH peak at *ca.* 3540 cm^{-1} in the latter was evidence that $ArgLi^+$ possessed a charge-solvated (canonical Arg) form, whereas the much lower intensity of this peak in $ArgNa^+$ indicated that the dominant species under their experimental conditions is a salt-bridged structure (zwitterionic Arg) [25]. These results were consistent with the relative energetics of the charge-solvated and salt-bridged forms of $ArgNa^+$ provided by their calculations [25]. Additional support for this assignment came from the IRMPD action spectrum of $ArgNa^+$ in the fingerprint region [26]. In this work, the lack of a strong C=O stretch near 1750 cm^{-1} in the spectrum of $ArgNa^+$ was taken as evidence of a salt-bridged structure, though they could not rule out the possibility of a minor component of charge-solvated $ArgNa^+$ in their gas-phase population based on a small peak in this region.

In contrast to the spectrum for $ArgNa^+$, the infrared action spectrum for $CavNa^+$ in the X–H region (Fig. 3c) shows a strong free OH stretch, similar to the spectra of $CavH^+$ (Fig. 3a) and $CavLi^+$ (Fig. 3b, see below) and quite similar to the non-zwitterionic $ArgLi^+$ spectrum from Bush et al. [25]. This spectrum suggests that the dominant species is charge solvated with a smaller contribution from additional structures that could be salt bridged. Fig. S5 of Supporting Information shows a comparison between the experimental IRMPD spectrum in the X–H region and calculated harmonic spectra for the lowest lying conformer of each isomer (**a–e**). As with $CavH^+$, the experimental spectrum for $CavNa^+$ (Fig. S5a) matches well with calculated spectra of the low-lying charge-solvated structures in Figs. S5b and S5c, but not with the calculated spectra for the salt-bridged isomers (Fig. S5d–f), in which the free OH stretching band is absent.

Additional spectroscopic evidence of the canonical structure of $CavNa^+$ is found in the fingerprint region IRMPD action spectrum of $CavNa^+$ obtained on the FEL/ion trap setup at FELIX (Fig. 3c). The strong peak at 1740 cm^{-1}, which is also present in the spectra of $CavH^+$ (Fig. 3a) and $CavK^+$ (Fig. 3d-1, from *m/z* 198, see below), corresponds to the C=O stretching frequency of the COOH group and is indicative of a charge-solvated structure. As with the X–H, spectrum, the fingerprint spectrum for $CavNa^+$ is in much better agreement with the $ArgLi^+$ (charge-solvated) fingerprint spectrum of Forbes et al. than with the spectrum of $ArgNa^+$ (salt-bridge) [26]. Fig. S5 also shows the experimental spectrum along with calculated spectra for the lowest-energy conformers of the canonical and zwitterionic isomers. The strong peak near 1740 cm^{-1} matches well with C=O carbonyl stretch in the charge-solvated structures. All of the low-lying can1 and can2 conformers that we found in our conformational search have this C=O stretch near 1740 cm^{-1}. In contrast, in the salt-bridged structures, the carbonyl COOH moiety is replaced with a carboxylate group and the asymmetric C=O stretching frequency is predicted to be red-shifted to between 1630 and 1660 cm^{-1} depending on the conformer. The spectrum in the fingerprint region is quite complex, and no single conformer of any isomer was found to match all of the peaks. There are seven low-lying charge-solvated conformers within 20 kJ/mol of the global minimum structure with similar but slightly different vibrational spectra. Based on the C=O peak near 1740 cm^{-1}, this spectrum confirms that the dominant gas-phase is charge-solvated, in contrast to the salt-bridged structure for $ArgNa^+$.

In their IRMPD study of $LysM^+$, Bush et al. showed that both the fingerprint and X–H IR spectra for $LysNa^+$ indicated that it is a charge- solvated structure. Their calculations at the B3LYP/6-31++G(d,p) level of theory predict that the difference between the lowest free-energy zwitterion and non-zwitterion for $LysNa^+$ is 3.4 kJ/mol at 298 K. We re-optimized their two lowest free-energy structures at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory to provide a consistent comparison with $ArgNa^+$ and $CavNa^+$. At this level, the difference in 298 K free energy is 9.2 kJ/mol, in reasonable agreement with the earlier study. Fig. 5 shows the lowest free-energy salt-bridged and charge-solvated structures for $XxxNa^+$ (Xxx = Arg, Cav, and Lys) and gives an indication of the types of structures afforded by the differing side chain compositions. Figs. 6–8 show similar structures for $XxxLi^+$, $XxxK^+$, and $XxxCs^+$ and show the effect of alkali metal size on the binding motifs for the different amino acids. For Na^+, the salt-bridged structures are similar for all three amino acids with the sodium cation interacting with the carboxylate oxygen atoms and the "protonated" amino terminus interacting with the side chain. For Arg, the side chain is actually more basic than the amino terminus so it favors a zw1 structure in which the extra proton is on the side chain. Since Cav and Lys are less basic, the extra proton resides on the amino terminus and the zwitterionic structures are not as stabilized by the salt bridge. An examination of the lowest free- energy conformers for the charge-solvated structures shows that $ArgNa^+$ and $LysNa^+$ adopt similar structures with the alkali metal interacting with the carbonyl oxygen atom and a nitrogen atom on the side chain, whereas $CavNa^+$ adopts a very different structure in which the sodium ion interacts with four heteroatoms, vastly stabilizing the charge-solvated form. Consequently, the differences in free energy between the salt-bridged and charge-solvated isomers, −13.5 kJ/mol for $ArgNa^+$, 15.3 kJ/mol for $CavNa^+$, and 9.2 kJ/mol for $LysNa^+$, can be rationalized as a combination of the differences in the basicity of the side chains in the salt-bridged structures and the differences in binding motif for the sodium ion and in hydrogen bonding motifs of the amino acids in the charge- solvated isomers.

Given that $CavNa^+$ and $CavH^+$ prefer charge- solvated structures, we would predict that $CavLi^+$ should also be charge solvated. The lowest-energy conformer for $CavLi^+$ is a can1 isomer (**a**) as shown in Fig. S6. The lithium ion is interacting strongly with the carbonyl oxygen atom, the amino terminal nitrogen atom, and one of the η-nitrogen atoms in the side chain. The lowest energy can2 isomer (**b**) lies only 3.8 kJ/mol higher in energy and involves coordination of the lithium ion to the carbonyl oxygen, the amino terminal nitrogen atom and the ε-nitrogen atom of the side chain. Our conformational search located 41 additional charge- solvated conformers within 40 kJ/mol of the global minimum structure, the lowest six of which are shown in Fig. S7 along with three additional zwitterionic conformers. A full tabulation of the relative free energies of all 307 salt-bridged and charge-solvated conformers for $CavLi^+$ is given in Table S5. The lowest-energy zwitterionic structure is of type zw1 (**c**) and involves lithium ion coordination to one of the carboxylate oxygen atoms and the amino terminal nitrogen atom. The other carboxylate oxygen interacts strongly with the protonated side chain forming a Li^+–CO_2^-–NH^+ salt-bridged isomer. This isomer lies 40.6 kJ/mol above the global minimum canonical structure and is shown in Fig. S6. It is analogous to structure G of Bush et al. for $ArgLi^+$, which is one of their three iso-energetic zwitterions at 410 K lying 12 kJ/mol above their lowest energy charge-solvated isomer [25]. The lowest energy zw3 (**e**) isomer for $CavLi^+$ is nearly iso-energetic with the (**c**) isomer (+40.9 kJ/mol) and involves lithium complexation to the two carboxylate oxygen atoms. A Li^+–CO_2^-–NH_3^+ salt-bridged structure is formed with the protonated amino terminal group (structure E for $ArgLi^+$[25]). The zw2 isomer (**d**) is less stable than the other zwitterions with the lowest energy conformer lying 54.1 kJ/mol above the global minimum canonical structure (similar to structure F of Bush et al. except that the shared proton between the N-terminus and the side chain is formally on the amino terminus) [25].

We ran a complete conformational search for $ArgLi^+$, including can2 and zw3 isomers. Due to the small size of the lithium cation, the can2 and zw3 isomers are less stable than the can1 and zw1 isomers found by Bush et al. (structures A and G). At our level of the-

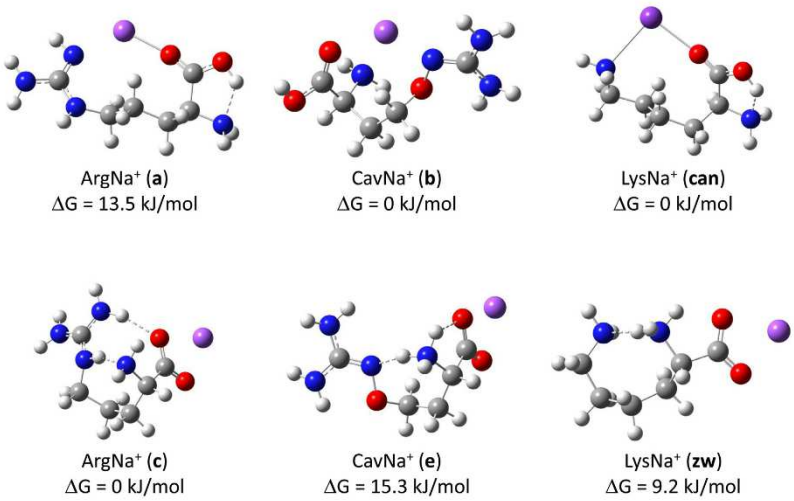


Fig. 5. Lowest free energy charge- solvated (top) and salt-bridged (bottom) isomers for $ArgNa^+$, $CavNa^+$, and $LysNa^+$. 298 K free energy differences calculated at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory.

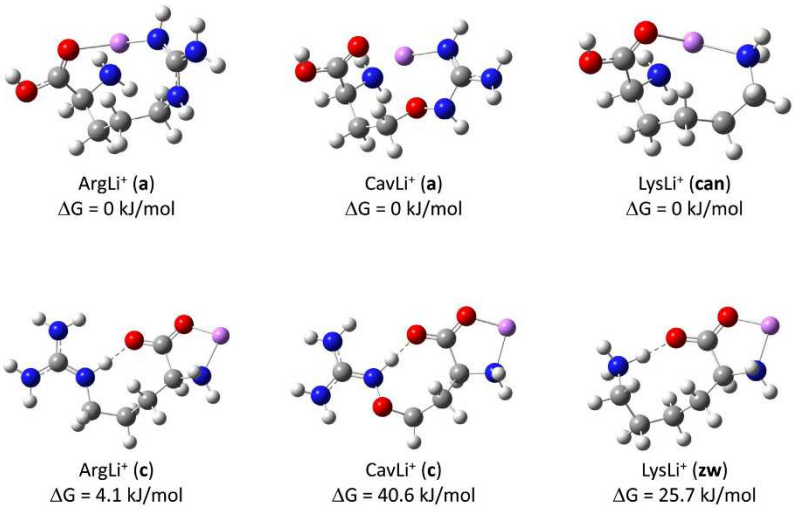


Fig. 6. Lowest free energy charge- solvated (top) and salt-bridged (bottom) isomers for $ArgLi^+$, $CavLi^+$, and $LysLi^+$. 298 K free energy differences calculated at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory.

ory, the difference between the global minimum charge-solvated structure (**a**) and the lowest-lying zwitterion (**c**) is 4.2 kJ/mol.

The IRMPD action spectrum of $CavLi^+$ in the X–H region was obtained at CLIO using the OPO/ICR setup (Fig. 3b). The strong free OH stretch at 3552 cm^{-1} provides experimental confirmation of the canonical nature of $CavLi^+$. Fig. S8 shows a comparison of the experimental spectrum to calculated spectra for low-lying canonical and zwitterionic species. As with $CavH^+$, the spectrum matches better with the calculated spectra of the low-lying canonical isomers than those of the zwitterion. Unfortunately, we were unable to obtain a fingerprint spectrum for $CavLi^+$ during our time at FELIX. We include in Fig. 3 the calculated fingerprint spectrum for the lowest energy canonical structure for $CavLi^+$ (**a**). This spectrum shows a strong carbonyl C=O stretch near 1730 cm^{-1} consistent with the charge-solvated ion containing a canonical Cav tautomer.

Fig. 6 shows a comparison of the lowest energy charge-solvated and salt-bridged isomers for $ArgLi^+$, $CavLi^+$, and $LysLi^+$. The binding motifs for all the amino acids are extremely similar due to the

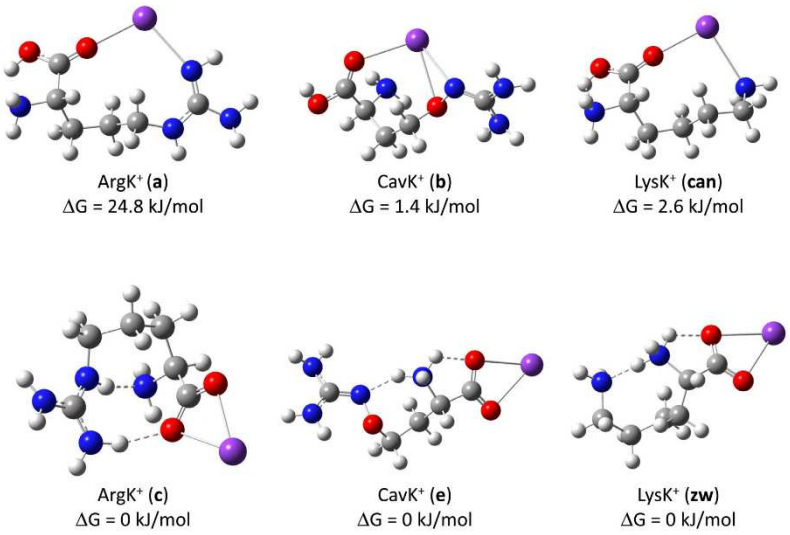


Fig. 7. Lowest free energy charge–solvated (top) and salt-bridged (bottom) isomers for $ArgK^+$, $CavK^+$, and $LysK^+$. 298 K free energy differences calculated at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory.

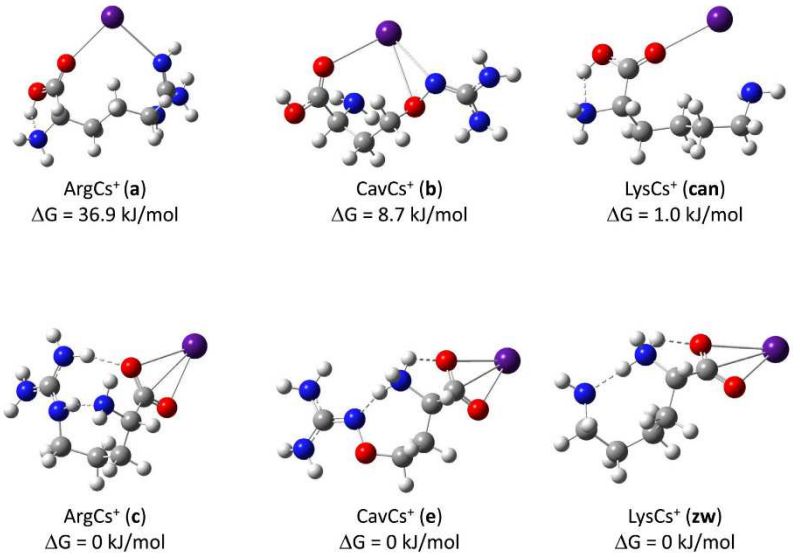


Fig. 8. Lowest free energy charge–solvated (top) and salt-bridged (bottom) isomers for $ArgCs^+$, $CavCs^+$, and $LysCs^+$. 298 K free energy differences calculated at the B3LYP/DEF2-TZVPP level of theory.

small size of the lithium cation. In the charge-solvated structures, the lithium ion has a strong interaction with the end of the guanidine/oxyguanidine/amine side chain, the amino terminus, and the carbonyl oxygen atom. In $CavLi^+$, there is an additional interaction with the oxygen atom (2.1 Å) in the side chain that stabilizes the charge-solvated form. The salt-bridged structures for all three amino acids are similar with a protonated side chain – carboxylate – Li^+ salt bridge and the lithium ion also interacting with

the amino terminus. Since the structures are similar, the stability of the salt-bridged structures should track the basicity of the side chain. As with the sodiated amino acids, the relative stability of the charge- solvated and salt-bridged isomers (+4.1 kJ/mol for ArgLi $^+$, +40.6 kJ/mol for CavLi$^+$, and +25.7 kJ/mol for LysLi$^+$) can be rationalized from the basicity of the side chains and the additional hydrogen/lithium ion binding site in Cav.

4.3. CavK$^+$

The lowest free-energy charge-solvated and salt-bridged conformers for CavK$^+$ are nearly isoenergetic at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level of theory. The lowest free-energy conformer is a zw3 salt-bridged isomer (**e**), with the lowest free-energy can2 charge-solvated (**b**) isomer lying only 1.4 kJ/mol higher in energy. The structures for these two isomers are shown in Fig. S9 and are most similar to Bush's structures E and A for ArgK$^+$, respectively [25]. The lowest free-energy can1 (**a**), zw1 (**c**), and zw2 (**d**) conformers are 26.1, 11.9 and 14.9 kJ/mol above the lowest free-energy conformer and are also shown in Fig. S9. Fig. S10 shows nine low-lying conformers for **5a–e**, and Table S6 gives relative free energy values for all 291 charge-solvated and salt-bridged isomers found in the conformational search. The lowest free-energy zwitterion has the potassium ion coordinating to the two carboxylate oxygen atoms and forming a K$^+$–CO_2^-–NH_3^+ salt bridge. This arrangement allows for an additional hydrogen bond between one of the hydrogens on the protonated amino group and the ε-N atom of the can2 tautomer side chain. The lowest energy charge-solvated structure has the K$^+$ interacting with the carbonyl oxygen, the amino terminus, and the oxygen atom and ε-N atom of the side chain. As with the other cationized Cav ions, the difference between salt-bridged and charge- solvated structures (1.9 kJ/mol) is much smaller than the analogous difference in ArgK$^+$ (28.1 kJ/mol) [26]. Our calculated structures for ArgK$^+$ are in excellent agreement with those of Forbes et al. [26] with the difference in free energy at 298 K equal to 24.8 kJ/mol between the B-type charge-solvated isomer and D-type salt- bridged isomer.

The theoretical results suggest that we should see features from a mixture of canonical and zwitterion structures in the infrared spectra of CavK$^+$. The action spectrum in the X—H region (Fig. 3d-1) looks noticeably different than that of CavH$^+$, CavLi$^+$, and CavNa$^+$. The intensity of the free O—H (3567 cm^{-1}) stretch has decreased, and the two N—H stretching peaks (3440 cm^{-1} and shoulder at 3550 cm^{-1}) are nearly identical in intensity. These differences point to a significant proportion of the zwitterion structure, though the presence of the free O—H stretch indicates that some portion of the population is charge-solvated. Fig. S11 of Supporting Information shows the experimental and the calculated spectra for the low-lying canonical and zwitterionic conformers for CavK$^+$. The experimental spectrum matches well with the spectrum for the lowest-energy zwitterion species for the two N—H stretching peaks at 3440 and 3550 cm^{-1} (Fig. S11b). The peak at 3377 cm^{-1} in the calculated spectrum that is missing in the experimental spectrum is an addition N—H stretch involving hydrogen atoms on the amino terminal nitrogen involved in strong hydrogen bonding with the side chain. The calculated spectrum for the lowest free-energy charge-solvated isomer shown in Fig. S11c matches the observed spectrum as well.

Additional evidence for a mixture of isomers is seen in the fingerprint spectrum obtained at FELIX shown in Fig. 3. As mentioned in the Experimental Section, in the ion trap at FELIX, CavK$^+$ dissociates to give fragments with m/z 171 (minor channel, loss of CO_2) and 198 (major channel, loss of NH_3). Fig. 3d-2 and d-3 shows the spectra for these individual ion loss channels. Clearly, the two spectra are quite different from each other and strongly suggest the presence of a mixture in the ion population. Fig. S11 shows the two experimental spectra and the calculated spectra for low-lying zwitterionic and canonical structures in the fingerprint region. The experimental spectrum for the loss of NH_3 channel (S11a-2) has a strong peak at 1745 cm^{-1}, which is suggestive of a charge-solvated structure. This peak is diminished in the experimental spectrum for the loss of CO_2 channel (S11a-3). The calculated spectra for the charge-solvated isomers (Fig. S11c and f) have this strong peak at 1750 cm^{-1} that is missing in the calculated spectra for the zwitterions (Fig. S11b, d, and e). We therefore assign the loss of NH_3 to a population of charge-solvated isomers and the loss of CO_2 to a population of zwitterionic isomers.

4.4. CavCs$^+$

Calculations on CavCs$^+$ required the use of an effective core potential basis set. Initial structures were taken from the CavNa$^+$ GMMX search by substituting Cs$^+$ for Na$^+$ in our Gaussian input file generator. An initial screening was done at the HF/LANL2DZ level and then geometries and frequencies for unique conformers of each isomer were calculated at the B3LYP/LANL2DZ level. Table S7 lists the relative free energy of all 303 charge-solvated and salt-bridged structures at the B3LYP/LANL2DZ level of theory. The lowest free-energy conformer at this level is a salt-bridged isomer of type zw3, and the lowest lying charge- solvated structure lies 42.6 kJ/mol higher in free energy. At the request of a reviewer, we re-optimized the 15 lowest energy CavCs$^+$ conformers and 13 additional CavCs$^+$ conformers from the LANL2DZ study using the DEF2-TZVPP basis set. The relative ordering between the three zwitterion types is the same at this level (zw3 < zw1 < zw2), as is the relative ordering of the two charge-solvated species (can2 < can1). The main difference between the two basis sets is the stability of the charge-solvated isomers relative to the salt-bridged isomers. The lowest energy DEF2-TZVPP charge-solvated structure has a very significant interaction of the cesium ion with the carbonyl oxygen atom (3.03 Å) and *both* the oxygen atom (3.16 Å) and ε-nitrogen atom (3.11 Å) of the side chain, whereas the lowest energy LANL2DZ charge-solvated isomer has decreased interactions between the cesium ion and the side chain atoms (3.49 Å and 3.21 Å, respectively). The difference in free energy between the salt-bridged and charge-solvated isomers drops from 42.6 kJ/mol to 8.7 kJ/mol with the DEF2-TZVPP basis set. Table S8 lists the relative free energies of the 28 CavCs$^+$ conformers at the DEF2-TZVPP level of theory. All free energy values for CavCs$^+$ presented in the paper except for Table S7 are at the DEF2- TZVPP level of theory.

The lowest free-energy conformer for CavCs$^+$ is a salt bridge of type zw3 as shown in Fig. S12. This structure is analogous to Forbes' structure E, except that the proton is on the amino terminus rather than on the side chain [26]. Low lying zw1 and zw2 isomers were located 12.7 and 15.4 kJ/mol above the zw3 isomer and are also shown in Fig. S12. The lowest lying charge-solvated ions with can2 and cav1 tautomers lie 8.7 kJ/mol and 37.8 kJ/mol above the zw3 global minimum structure. Fig. S13 gives structures and relative free energies for 12 additional low-energy conformers. The lowest free-energy conformers for ArgCs$^+$ at the B3LYP/DEF2-TZVPP level are identical with those found by Forbes. As with the other metals, the charge- solvated isomers for CavCs$^+$ are stabilized with respect to the salt-bridged isomer (8.7 kJ/mol) more than in ArgCs$^+$ (36.9 kJ/mol).

The X—H spectrum for CavCs$^+$ shown in Fig. 3e is consistent with a salt-bridged structure since the free OH stretch is missing and the NH_2 stretching bands are of a similar intensity. Fig. S14 of Supporting Information shows the experimental spectrum along with calculated spectra for zw3, zw1, and can2 isomers using scaled (0.965) B3LYP/DEF2-TZVPP frequencies. The calculated spectrum for the zw3 salt-bridged isomer (Fig. S14b) shows the two sets of symmetric and antisymmetric N—H stretching modes of the

η-NH_2 groups that agree reasonably well with the experimental spectrum. In the zw1 salt-bridged structure, one of the NH_2 groups is involved in a strong hydrogen bond with the carboxylate ion, which red-shifts the symmetric and anti-symmetric N—H stretches. This calculated spectrum (Fig. S14d) does not match as well with the experimental spectrum, nor does the calculated spectrum for the for the zw2-type isomer (Fig. S14e), though the experimental spectrum cannot rule out a mixture of different zwitterionic structures. The calculated spectrum for the lowest free-energy charge-solvated isomer (Fig. S14c) has a strong free OH stretch that is absent from the experimental spectrum strongly suggesting the lack of charge-solvated isomers in the gas-phase population.

Additional insight into the structure of $CavCs^+$ is provided in its fingerprint spectrum shown in the top trace of Fig. 3. The absence of a strong C═O stretching band near 1750 cm^{-1} indicates that the vast majority of the isomers of $CavCs^+$ are salt bridged. There is a small shoulder in the experimental spectrum that may indicate the presence of a small population of charge-solvated isomers. Fig. S15 shows the experimental spectrum and calculated spectra for the five different $CavCs^+$ isomers. The calculated spectra for the can1 and can2 charge-solvated isomers (Figs. S15c and f) both show a strong C═O stretching band that is absent from the experimental fingerprint spectrum, again suggesting that the gas-phase population does not contain charge-solvated isomers. The calculated spectra from the three different zwitterion classes match in the sense that they do not contain the free C═O stretch but, as with the $CavK^+$ spectra, no single calculated spectrum matches the experimental spectra. Given the relative energetics and the agreement with calculated spectra in the X—H region of the spectrum, we conclude that $CavCs^+$ is a salt-bridged structure.

5. Conclusions

Infrared spectra in the X—H and fingerprint regions of protonated and alkali-metallated canavanine were obtained using IRMPD. In contrast to arginine, which forms charge-solvated ions with Li^+ and salt-bridged structures for Na^+, K^+, Rb^+, and Cs^+, and lysine which forms charge- solvated structures for all alkali metal ions, $CavLi^+$ and $CavNa^+$ were found to be charge-solvated, $CavK^+$ was found to be a mixture of charge-solvated and salt-bridged isomers, and $CavCs^+$ was found to be mostly salt bridged. Canavanine has an oxygen atom substituted into its side chain which contributes to the difference in gas-phase structure. The electron withdrawing nature of the oxygen atom changes the charge distribution and severely decreases the overall basicity of the side chain relative to that of arginine. The decreased basicity of the side chain causes canavanine to form zwitterions with a protonated amino terminus rather than a protonated side chain for all alkali metals except Li^+. Thus, salt bridges of the form $M^+–CO_2^-–H_3N^+_{amino}$ are formed rather than the $M^+–CO_2^-–HN^+_{sidechain}$ structures found in metallated-Arg, allowing for differential hydrogen bonding schemes in the salt-bridged isomers. Similar $M^+–CO_2^-–H_3N^+_{amino}$ salt bridges are found for $LysM^+$, which form charge-solvated structures for M = Li, Na, K, and Cs.

The decreased basicity of the side chain is clearly not the only factor that affects which form Cav takes when complexed with alkali metals. The oxygen atom in the side chain of Cav can act as an additional hydrogen-bonding acceptor site as well as an additional alkali metal binding site and allows for enhanced stability of the charged-solvated structures of Cav complexes with the smaller alkali metals. Calculated structures for $CavLi^+$, $CavNa^+$, $CavK^+$, and $CavCs^+$ show significant interactions between the metal ion and the side-chain oxygen atom in the charge-solvated forms. The extra stability of these canonical forms combines with the decreased salt-bridged stability from the less basic side chain to shift the cross-over point for charge-solvated to salt-bridged forms from between Li^+ and Na^+ for $ArgM^+$ to between K^+ and Cs^+ for $CavM^+$.

Acknowledgements

Funding for this project was generously provided by the National Science Foundation (CHE0911244 and CHEM:1464763), the National Institutes of Health, (1R15GM116180-01), and NWO Chemical Sciences under VICI project no. 724.011.002 (FELIX).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijms.2017.08.009.

References

[1] L. Stryer, Biochemistry, 3rd ed., W. H. Freeman and Co., New York, 1988.
[2] R.A.J. O'Hair, J.H. Bowie, S. Gronert, Gas phase acidities of the α-amino acids, Int. J. Mass Spectrom. Ion Process. 117 (1992) 23.
[3] C.M. Jones, M. Bernier, E. Carson, K.E. Colyer, R. Metz, A. Pawlow, E. Wischow, I. Webb, E.J. Andriole, J.C. Poutsma, Gas- phase acidities of the 20 protein amino acids, Int. J. Mass Spectrom. 267 (2007) 54.
[4] C. Bleiholder, S. Suhai, B. Paizs, Revising the proton affinity scale of the naturally occurring α-amino acids, J. Am. Soc. Mass Spectrom. 17 (2006) 1275.
[5] S. Gronert, D.C. Simpson, K.M. Conner, A reevaluation of computed proton affinities for the common α-amino acids, J. Am. Soc. Mass Spectrom. 20 (2009) 2116.
[6] E. Kassab, J. Langlet, E. Evleth, Y. Akacem, Theoretical study of solvent effect on intramolecular proton transfer of glycine, J. Mol. Struct. (THEOCHEM) 531 (2000) 267.
[7] Z. Wu, C. Fenselau, Proton affinity of arginine measured by the kinetic approach, Rapid Comm. Mass Spectrom. 6 (1992) 403.
[8] G. Bouchoux, S. Desaphy, S. Bourcier, C. Malosse, R. Bimbong, Gas phase protonation thermochemistry of arginine, J. Phys. Chem. B 112 (2008) 3410.
[9] S. Ling, W. Yu, Z. Huang, Z. Lin, M. Haranczyk, M. Gutowski, Gaseous arginine conformers and their unique intramolecular interactions, J. Phys. Chem. A 110 (2006) 12282.
[10] C.J. Chapo, J.B. Paul, R.A. Provencal, K. Roth, R.J. Saykallay, Is arginine zwitterionic or neutral in the gas phase? Results from IR cavity ringdown spectroscopy, J. Am. Chem. Soc. 120 (1998) 12956.
[11] J. Jensen, M. Gordon, On the number of water molecules necessary to stabilize the glycine zwitterion, J. Am. Chem. Soc. 113 (1991) 7917.
[12] S. Xu, J.M. Nilles, K.H. Bowin Jr., Zwitterion formation in hydrated amino acid, dipole bound anions: how many water molecules are required? J. Chem. Phys. 119 (2003) 10696.
[13] E.G. Diken, N.I. Hammer, M.A. Johnson, Preparation and photoelectron spectrum of the glycine molecular anion: assignment to a dipole-bound electron species with a high-dipole moment, nonzwitterionic form of the neutral core, J. Chem. Phys. 120 (2004) 9899.
[14] E.G. Diken, J.M. Headrick, M.A. Johnson, Photoelectron spectroscopy of the $[glycine(H_2O)_{1,2}]^-$ clusters: sequential hydration shifts and observation of isomers, J. Chem. Phys. 112 (2005) 224317.
[15] C.M. Aikens, M.S. Gordon, Incremental solvation of nonionized and zwitterionic glycine, J. Am. Chem. Soc. 128 (2006) 12835.
[16] S.-W. Park, S. Im, S. Lee, C. Desfrancois, Structure and stability of glycine-$(H_2O)_2$ cluster and anion: zwitterion vs. canonical glycine, Int. J. Quantum Chem. 107 (2007) 1316.
[17] T. Takayanagi, T. Yoshikawa, A. Kakizaki, M. Shiga, M. Tachikawa, Molecular dynamics simulations of small glycine-$(H_2O)_n$ (n = 2–7) clusters on semiempirical PM6 potential energy surfaces, J. Mol. Struct. (THEOCHEM) 867 (2008) 29.
[18] J.-Y. Kim, S. Im, B. Kim, C. Desfrancois, S. Lee, Structures and energetics of Gly-(H2O)5: thermodynamic and kinetic stabilities, Chem. Phys. Lett. 451 (2008) 198.
[19] P.K. Sahu, S.-L. Lee, Effect of microsolvation on zwitterionic glycine: an ab initio and density functional theory study, J. Mol. Model. 14 (2008) 385.
[20] K.T. Lee, K.Y. Han, I. Oh, S.K. Kim, Barrierless pathways in the neutral-zwitterion transition of amino acid: glycine-$(H_2O)_9$, Chem. Phys. Lett. 495 (2010) 14.
[21] X.-J. Meng, H.-L. Zhao, X.S. Ju, Influences of n (2–5) water molecules on the proton transfer in hydrated glycine complexes, Comput. Theor. Chem. 1001 (2012) 26.
[22] R. Peïrez de Tudela, D. Marx, Water-induced zwitterionization of glycine: stabilization mechanism and spectral signatures, J. Phys. Chem. Lett. 7 (2016) 5137.
[23] J.-Y. Kim, D.-S. Ahn, S.-W. Park, S. Lee, Gas phase hydration of amino acids and dipeptides: effects on the relative stability of zwitterion vs. canonical conformers, RSC Adv. 4 (2014) 16352.

[24] M.F. Bush, J.S. Prell, R.J. Saykallay, E.R. Williams, One water molecule stabilizes the cationized arginine zwitterion, J. Am. Chem. Soc. 129 (2007) 13544.
[25] M.F. Bush, J.T. O'Brien, J.S. Prell, R.J. Saykallay, E.R. Willliams, Infrared spectroscopy of cationized arginine in the gas phase: direct evidence for the transition from nonzwitterionic to zwitterionic structure, J. Am. Chem. Soc. 129 (2007) 1612.
[26] M.W. Forbes, M.F. Bush, N.C. Polfer, J. Oomens, R.C. Dunbar, E.R. Williams, R.A. Jockusch, Infrared spectroscopy of arginine cation complexes: direct observation of gas-phase zwitterions, J. Phys. Chem. A 111 (2007) 11759.
[27] R.R. Julian, R. Hodyss, J.L. Beauchamp, Salt bridge stabilization of charged zwitterionic arginine aggregates in the gas phase, J. Am. Chem. Soc. 123 (2001) 3577.
[28] O. Gellar, C. Lifshitz, Gas phase H/D exchange of protonated arginine monomers and dimers, J. Phys. Chem. A 107 (2003) 5654.
[29] W.D. Price, R.A. Jockusch, E.R. Williams, Is arginine a zwitterion in the gas phase? J. Am. Chem. Soc. 119 (1997) 11988.
[30] R. Wu, R.A. Marta, J.K. Martens, K.R. Eldridge, T.B. McMahon, Experimental and theoretical investigation of the proton-bound dimer of lysine, J. Am. Soc. Mass Spectrom. 22 (2011) 1651.
[31] T. Wyttenbach, M. Witt, M.T. Bowers, On the stability of amino acid zwitterions in the gas phase: the influence of derivatization, proton affinity, and alkali ion addition, J. Am. Chem. Soc. 122 (2000) 3458.
[32] R.A. Jockusch, W.D. Price, E.R. WIlliams, Structure of cationized arginine (Arg·M+, M = H, Li, Na, K, Rb, and Cs) in the gas phase: further evidence for zwitterionic arginine, J. Phys. Chem. A 103 (1999) 9266.
[33] B.A. Cerda, C. Wesdemiotis, Zwitterionic vs. charge-solvated structures in the binding of arginine to alkali metal ions in the gas phase, Analyst 125 (2000) 657.
[34] J.R. Eyler, Infrared multiphoton dissociation spectroscopy of ions in penning traps, Mass Spectrom. Rev. 28 (2009) 448.
[35] N.C. Polfer, J. Oomens, Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance, Mass Spectrom. Rev. 28 (2009) 468.
[36] T.D. Fridgen, Infrared consequence spectroscopy of gaseous protonated and metal ion cationized complexes, Mass Spectrom. Rev. 28 (2009) 586.
[37] J. Martens, J. Grzetic, G. Berden, J. Oomens, Structural identification of electron transfer dissociation products in mass spectrometry using infrared ion spectroscopy, Nat. Commun. 7 (2016) 11754.
[38] J. Lemaire, P. Boissel, M. Heninger, G. Mauclaire, G. Bellec, H. Mestdagh, A. Simon, S. LeCaer, J.M. Ortega, F. Glotin, P. Maitre, Gas phase infrared spectroscopy of selectively prepared ions, Phys. Rev. Lett. 89 (2002) 273002.
[39] P. Maitre, S. LeCaer, A. Simon, W. Jones, J. Lemaire, H. Mestdagh, M. Heninger, G. Mauclaire, P. Boissel, R. Prazeres, F. Glotin, J.M. Ortega, Ultrasensitive spectroscopy of ionic reactive intermediates in the gas phase performed with the first coupling of an IR FEL with an FTICR-MS, Nucl. Instrum. Methods A 507 (2003) 541.
[40] J.J. Valle, J.R. Eyler, J. Oomens, D.T. Moore, A.F.G. van der Meer, G. von Helden, G. Meijer, C.L. Hendrickson, A.G. Marshall, G.T. Blakney, Free electron laser-Fourier transform ion cyclotron resonance mass spectrometry facility for obtaining infrared multiphoton dissociation spectra of gaseous ions, Rev. Sci. Instrum. 76 (2005) 023103.
[41] A. Kamariotis, O.V. Boyarkin, S.R. Mercier, R.D. Beck, M.F. Bush, E.R. Williams, T.R. Rizzo, Infrared spectroscopy of hydrated amino acids in the gas phase: protonated and lithiated valine, J. Am. Chem. Soc. 2006 (2006) 905.
[42] M.B. Burt, S.G.A. Decker, C.G. Atkins, M. Roswell, A. Peremans, T.D. Fridgen, Structures of bare and hydrated [Pb(AminoAcid-H)]+ complexes using infrared multiple photon dissociation spectroscopy, J. Phys. Chem. B 115 (2011) 11506.
[43] J. Schmidt, M.M. Meyer, I. Spector, S.R. Kass, Infrared multiphoton dissociation spectroscopy study of protonated p-aminobenzoic acid: does electrospray ionization afford the amino- or carboxy-protonated ion? J. Phys. Chem. A 115 (2011) 7625.
[44] K. Gulyuz, C.N. Stedwell, D. Wang, N.C. Polfer, Hybrid quadrupole mass filter/quadrupole ion trap/time-of-flight-mass spectrometer for infrared multiple photon dissociation spectroscopy of mass-selected ions, Rev. Sci. Instrum. 82 (2011) 054101.
[45] J.K. Martens, J. Grzetic, G. Berden, J. Oomens, Gas-phase conformations of small polyprolines and their fragment ions by IRMPD spectroscopy, Int. J. Mass Spectrom. 377 (2015) 179.
[46] C. Kapota, J. Lemaire, P. Maitre, G. Ohanessian, Vibrational signature of charge solvation vs salt bridge isomers of sodiated amino acids in the gas phase, J. Am. Chem. Soc. 126 (2004) 1836.
[47] N.C. Polfer, B. Paizs, L.C. Snoek, I. Compagnon, S. Suhai, G. Meijer, G. Von Helden, J. Oomens, Infrared fingerprint spectroscopy and theoretical studies of potassium ion tagged amino acids and peptides in the gas phase, J. Am. Chem. Soc. 127 (2005) 8571.
[48] N.C. Polfer, J. Oomens, R.C. Dunbar, IRMPD spectroscopy of metal-ion/tryptophan complexes, Phys. Chem. Chem. Phys. 8 (2006) 2744.
[49] N.C. Polfer, J. Oomens, D.T. Moore, G. Von Helden, G. Meijer, R.C. Dunbar, Infrared spectroscopy of phenylalanine Ag(I) and Zn(II) complexes in the gas phase, J. Am. Chem. Soc. 128 (2006) 517.
[50] R. Wu, T.B. McMahon, Stabilization of the zwitterionic structure of proline by an alkylammonium ion in the gas phase, Angew. Chem. Int. Ed. 46 (2007) 3668.
[51] R. Wu, T.B. McMahon, Infrared multiple photon dissociation spectra of proline and glycine proton-bound homodimers. Evidence for zwitterionic structure, J. Am. Chem. Soc. 129 (2007) 4864.
[52] M.F. Bush, M.W. Forbes, R.A. Jockusch, J. Oomens, N.C. Polfer, R.J. Saykallay, E.R. Williams, Infrared spectroscopy of cationized lysine and ε-N-methyllysine in the gas phase: effects of alkali-metal ion size and proton affinity on zwitterion stability, J. Phys. Chem. A 111 (2007) 7753.
[53] N.C. Polfer, R.C. Dunbar, J. Oomens, Observation of zwitterion formation in the gas-phase H/D-exchange with CH3OD: solution-phase structures in the gas phase, J. Am. Soc. Mass Spectrom. 18 (2007) 512.
[54] R.C. Dunbar, N.C. Polfer, J. Oomens, Gas-phase zwitterion stabilization by a metal dication, J. Am. Chem. Soc. 129 (2007) 14562.
[55] R. Wu, T.B. McMahon, IRMPD spectra of Gly·NH_4^+ and proton-bound betaine dimer: evidence for the smallest gas-phase zwitterionic structures, J. Mass Spectrom. 43 (2008) 1641.
[56] J.T. O'Brien, J.S. Prell, J.D. Steill, J. Oomens, E.R. Williams, Interactions of mono- and divalent metal ions with aspartic and glutamic acid investigated with IR photodissociation spectroscopy and theory, J. Phys. Chem. A 112 (2008) 10823.
[57] M.F. Bush, J. Oomens, R.J. Saykallay, E.R. Williams, Alkali metal ion binding to glutamine and glutamine derivatives investigated by infrared action spectroscopy and theory, J. Phys. Chem. A 112 (2008) 8578.
[58] M.F. Bush, J. Oomens, R.J. Saykallay, E.R. Williams, Effects of alkaline earth metal ion complexation on amino acid zwitterion stability: results from infrared action spectroscopy, J. Am. Chem. Soc. 130 (2008) 6463.
[59] A.L. Heaton, V.N. Bowman, J. Oomens, J.D. Steill, P.B. Armentrout, Infrared multiple photon dissociation spectroscopy of cationized asparagine: effects of metal cation size on gas-phase conformation, J. Phys. Chem. A 113 (2009) 5519.
[60] R.C. Dunbar, J.D. Steill, N.C. Polfer, J. Oomens, Dimeric complexes of tryptophan with M^{2+} metal ions, J. Phys. Chem. A 113 (2009) 845.
[61] M.F. Bush, J. Oomens, E.R. Williams, Proton affinity and zwitterion stability: new results from infrared spectroscopy and theory of cationized lysine and analogues in the gas phase, J. Phys. Chem. A 113 (2009) 431.
[62] M.K. Drayss, P.B. Armentrout, J. Oomens, M. Schaefer, IR spectroscopy of cationized aliphatic amino acids: stability of charge-solvated structure increases with metal cation size, Int. J. Mass Spectrom. 297 (2010) 18.
[63] M. Citir, E.M.S. Stennett, J. Oomens, J.D. Steill, M.T. Rodgers, P.B. Armentrout, Infrared multiple photon dissociation spectroscopy of cationized cysteine: effects of metal cation size on gas-phase conformation, Int. J. Mass Spectrom. 297 (2010) 9.
[64] D.R. Carl, T.E. Cooper, J. Oomens, J.D. Steill, P.B. Armentrout, Infrared multiple photon dissociation spectroscopy of cationized methionine: effects of alkali-metal cation size on gas-phase conformation, Phys. Chem. Chem. Phys. 12 (2010) 3384.
[65] J.T. O'Brien, J.S. Prell, G. Berden, J. Oomens, E.R. Williams, Effects of anions on the zwitterion stability of Glu, His and Arg investigated by IRMPD spectroscopy and theory, Int. J. Mass Spectrom. 297 (2010) 116.
[66] T.E. Hofstetter, C. Howder, G. Berden, J. Oomens, P.B. Armentrout, Structural elucidation of biological and toxicological complexes: investigation of monomeric and dimeric complexes of histidine with multiple charged transition metal (Zn and Cd) cations using IR action spectroscopy, J. Phys. Chem. B 115 (2011) 12648.
[67] W.K. Mino, K. Gulyuz, D. Wang, C.N. Stedwell, N.C. Polfer, Gas- phase structure and dissociation chemistry of protonated tryptophan elucidated by infrared multiple- photon dissociation spectroscopy, Phys. Chem. Lett. 2 (2011) 299.
[68] M. Citir, C.S. Hinton, J. Oomens, J.D. Steill, P.B. Armentrout, Infrared multiple photon dissociation spectroscopy of protonated histidine and 4-phenylimidazole, Int. J. Mass Spectrom. 330–332 (2012) 6.
[69] M. Citir, C.S. Hinton, J. Oomens, J.D. Steill, P.B. Armentrout, Infrared multiple photon dissociation spectroscopy of cationized histidine: effects of metal cation size on gas-phase conformation, J. Phys. Chem. A 116 (2012) 1532.
[70] J. Schmidt, S.R. Kass, Zwitterion vs neutral structures of amino acids stabilized by a negatively charged site: infrared photodissociation and computations of proline-chloride anion, J. Phys. Chem. A 117 (2013) 4863.
[71] D. Scuderi, E. Bodo, B. Chiavarino, S. Fornarini, M.E. Crestoni, Amino acid oxidation: a combined study of cysteine oxo forms by IRMPD spectroscopy and simulations, Chem. Eur. J. 22 (2016) 17239.
[72] Y. Jami-Alahmadi, T.D. Fridgen, Structures and unimolecular chemistry of M(Pro2–H) + (M = Mg, Ca, Sr, Ba, Mn, Fe, Co, Ni, Cu, Zn) by IRMPD spectroscopy, SORI- CID, and theoretical studies, Phys. Chem. Chem. Phys. 18 (2016) 2023.
[73] R.A. Coates, C.P. McNary, G.C. Boles, G. Berden, J. Oomens, P.B. Armentrout, Structural characterization of gas-phase cysteine and cysteine methyl ester complexes with zinc and cadmium dications by infrared multiple photon dissociation spectroscopy, Phys. Chem. Chem. Phys. 17 (2015) 25799.
[74] K. Rajabi, T.D. Fridgen, Structures of aliphatic amino acid proton-bound dimers by infrared multiple photon dissociation spectroscopy in the 700–2000 cm^{-1} region, J. Phys. Chem. A 112 (2008) 23.
[75] C.G. Atkins, K. Rajabi, E.A.L. Gillis, T.D. Fridgen, Infrared multiple photon dissociation spectra of proton- and sodium ion-bound glycine dimers in the N-H and O-H stretching region, J. Phys. Chem. A 112 (2008) 10220.
[76] J.T. O'Brien, J.S. Prell, J.D. Steill, J. Oomens, E.R. Williams, Changes in binding motif of protonated heterodimers containing valine and amines

investigated using IRMPD spectroscopy between 800 and 3700 cm^{-1} and theory, J. Am. Chem. Soc. 131 (2009) 3905.
[77] S. Jaeqx, J. Oomens, A.M. Rijs, Gas-phase salt bridge interactions between glutamic acid and arginine, Phys. Chem. Chem. Phys. 15 (2013) 16341.
[78] M. Blom, I. Compagnon, N.C. Polfer, G. Von Helden, G. Meijer, S. Suhai, B. Paizs, J. Oomens, Stepwise solvation of an amino acid: the appearance of zwitterionic structures, J. Phys. Chem. A 111 (2007) 7309.
[79] C.G. Atkins, L. Banu, M. Roswell, V. Blagojevic, D.K. Bohme, T.D. Fridgen, Structure of $[Pb(Gly-H)]^+$ and the monosolvated water and methanol solvated species by infrared multiple-photon dissociation spectroscopy, energy-resolved collision-induced dissociation, and electronic structure calculations, J. Phys. Chem. B 113 (2009) 14457.
[80] J.S. Prell, T.C. Correra, T.M. Chang, J.A. Biles, E.R. Williams, Entropy drives an attached water molecule from the C- to N-terminus on protonated proline, J. Am. Chem. Soc. 132 (2010) 14733.
[81] J.S. Prell, T.M. Chang, J.T. O'Brien, E.R. Williams, Hydration isomers of protonated phenylalanine and derivatives: relative stabilities from infrared photodissociation, J. Am. Chem. Soc. 132 (2010) 7811.
[82] A. Gholami, T.D. Fridgen, Structures and unimolecular reactivity of gas-phase $[Zn(Proline-H)]^+$ and $[Zn(Proline-H)(H2O)]^+$, J. Phys. Chem. B 117 (2013) 8447.
[83] M.B. Moghaddam, T.D. Fridgen, IRMPD spectroscopic study of microsolvated $[Na(GlyAla)]^+$ and $[Ca(GlyAla-H)]^+$ and the blue shifting of the hydrogen-bonded amide stretch with each water addition, J. Phys. Chem. B 117 (2013) 6157.
[84] M.B. Burt, T.D. Fridgen, Gas- phase structures of Pb^{2+}-cationized phenylalanine and glutamic acid determined by infrared multiple photon dissociation spectroscopy and computational chemistry, J. Phys. Chem. A 117 (2013) 1283.
[85] G.S. Gorman, J.P. Spier, C.A. Turner, I.J. Amster, Proton affinities of the 20 common α-amino acids, J. Am. Chem. Soc. 114 (1992) 3986.
[86] Z. Wu, C. Fenselau, Gas phase basicities and proton affinities of lysine and histidine measured from the dissociation of proton-bound dimers, Rapid Commun. Mass Spectrom. 8 (1994) 777.
[87] O.E. Schroeder, E.J. Andriole, K.L. Carver, J.C. Poutsma, The proton affinity of lysine analogs using the extended kinetic method, J. Phys. Chem. A 108 (2004) 326.
[88] G. Bouchoux, D. Buisson, C. Colas, M. Sablier, Protonation thermochemistry of α-amino acids bearing a basic residue, Eur. J. Mass Spectrom. 10 (2004) 977.
[89] E.A. Bell, Non- protein amino acids in plants, Encylc. Plant Phys. 8 (1980) 403.
[90] C.S. Evans, E.A. Bell, Uncommon amino acids in 64 species of caesalpinieae, Phytochemistry 17 (1978) 1127.
[91] S.L.N. Rao, P.R. Adiga, P.S. Sarma, The isolation and characterization of beta-N-oxalyl-L-alpha,beta-diaminopropionic acid: a neurotoxin from the seed of Lathrus sativus, Biochemistry 47 (1964) 432.
[92] G.A. Rosenthal, The biological effects and mode of action of L-canavanine, a structural analogue of L-arginine, Q. Rev. Biol. 52 (1977) 155.
[93] A. Boyar, R.E. Marsh, L- Canavanine, a paradigm for the structures of substituted guanidines, J. Am. Chem. Soc. 104 (1982) 1995.
[94] G.A. Rosenthal, D.H. Janzen, Avoidance of non-protein amino acid incorporation into protein by the seed predator *Caryede brasiliensis* (Bruchidae), J. Chem. Ecol. 9 (1983) 1353.
[95] M.J. Carrington, D. Fornieri, C.P. Bartlet, G. Heale, C.I. Levene, Biochemical and ultrastructural study of the effects of proline analogues on collagen synthesis in 3T6 fibroblasts, J. Pharm. Pharmacol. 37 (1985) 794.
[96] L. Hue, H.S. Sherratt, Inhibition of gluconeogenesis by hypoglycin in the rat: evidence for inhibition of glucose-6-phosphatase in vivo, Biochemistry 240 (1986) 765.
[97] P.S. Spencer, P.B. Nunn, J. Hugon, Guam amyotropic lateral sclerosis-parkinsonism-dementia linked to a plant excitant neurotoxin, Science 237 (1987) 517.
[98] A.C. Ludolph, J. Hugon, M.P. Dwivedi, H.H. Schaumburg, P.S. Spencer, Studies on the aetiology and pathogenesis of motor neuron diseases: 1. Lathyrism: clinical findings in established cases, Brain 110 (1987) 149.
[99] T.D. Copeland, E.M. Wondrak, J. Tozser, M.M. Roberts, S. Oroszian, Substitution of proline with pip at the scissile bond converts a peptide substrate of HIV proteinase into a selective inhibitor, Biochem. Biophys. Res. Commun. 169 (1990) 310.
[100] Y. Ikeda, K. Tanaka, Selective inactivation of various Acyl-CoA dehydrogenases by (methylenecyclopropyl)acetyl-CoA, Biochim. Biophys. Acta 1038 (1990) 216.
[101] K. Tanaka, Y. Ikeda, Hypoglycin and Jamaican vomiting sickness, Prog. Clin. Biol. Res. 321 (1990) 167.
[102] G.A. Rosenthal, The biochemical basis for the deleterious effects of L-canavanine, Phytochemistry 30 (1991) 1055.
[103] M. Lai, H. Liu, New evidence supporting a radical mechanism of the inactivation of general acyl-CoA dehydrogenase by a metabolite of hypoglycin, J. Am. Chem. Soc. 114 (1992) 3160.
[104] N. Carvajal, C. Torres, E. Uribe, M. Salas, Interaction of arginase with metal ions: studies of the enzyme from human liver and comparison with other arginases, Comp. Biochem. Physiol. 112 (1995) 153.
[105] T.J. Deming, M.J. Fournier, T.L. Mason, D.A. Tirrell, Structural modification of a periodic polypeptide through biosynthetic replacement of proline with aze, Macromolecules 29 (1996) 1442.
[106] C. Melangeli, G.A. Rosenthal, D.L. Dalman, The biochemical basis for L-canavanine tolerance by the tobacco budworm *Heliothis virescens* (Noctuidae), Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 2255.
[107] A.F. Kuntz, A.W. Boynton, G.A. David, K.E. Colyer, J.C. Poutsma, Proton affinities of proline analogs using the kinetic method with full entropy analysis, J. Am. Soc. Mass Spectrom. 13 (2002) 72.
[108] I. Webb, C. Muetterties, C.B. Platner, J.C. Poutsma, Gas- phase acidities of lysine homologues and proline analogs from the extended kinetic method, Int. J. Mass Spectrom. 316–318 (2012) 126.
[109] M.M. Raulfs, L. Breci, M. Bernier, O. Hamdy, A. Janiga, V.H. Wysocki, J.C. Poutsma, Investigations of the mechanism of the 'Proline Effect' in mass spectrometry peptide fragmentation experiments: the pipecolic acid effect, J. Am. Soc. Mass Spectrom. 25 (2014) 1705.
[110] C. Muetterties, A. Janiga, K.T. Huynh, M.G. Pisano, V.T. Tripp, D.D. Young, J.C. Poutsma, Gas- phase acid-base properties of homocysteine, homoserine, 5-mercaptonorvaline, and 5-hydroxynorvaline from the extended kinetic method, Int. J. Mass Spectrom. 369 (2014) 71.
[111] E.J. Andriole, K.E. Colyer, E. Cornell, J.C. Poutsma, Proton affinity of canavanine and canaline, oxy-analogs of arginine and ornithine, from the extended kinetic method, J. Phys. Chem. A 110 (2006) 11501.
[112] Y.-H. Ju, A.-W. Khaleel, A.E. Fazary, Guanidinium protonation equilibria of L-canavanine in different ionic media, Chem. Eng. Data 55 (2010) 3772.
[113] A.K. Bence, P.A. Crooks, The mechanism of L-canavanine cytotoxicity: arginyl tRNA synthetase as a novel target for anticancer drug discovery, J. Enzyme Inhib. Med. Chem. 18 (2003) 383.
[114] G.L. Igloi, E. Schiefermayr, Amino acid discrimination by arginyl-tRNA synthetases as revealed by an examination of natural specificity variants, FEBS J. 276 (2009) 1307.
[115] G.A. Rosenthal, D.L. Dahlman, D.H. Janzen, A novel means for dealing with L-canavanine, a toxic metabolite, Science 192 (1976) 256.
[116] A.K. Bence, D.R. Worthen, V.R. Adams, P.A. Crooks, The antiproliferative and immunotoxic effects of L-canavanine and L-canaline, Anti-Cancer Drug 13 (2002) 313.
[117] A.D.R. Nurcahyanti, M. Wink, Cytotoxic potentiation of vinblastine and paclitaxel by L-canavanine in human cervical cancer and hepatocellular carcinoma cells, Phytomedicine 22 (2015) 1232.
[118] A.D.R. Nurcahyanti, M. Wink, L- Canavanine potentiates cytotoxicity of chemotherapeutic drugs in human breast cancer cells, Anticancer Agents Med. Chem. 17 (2017) 206.
[119] B. Vynnytska-Myronovska, Y. Bobak, Y. Garbe, C. Dittfield, O. Stasyk, L.A. Kunz-Schughart, Single amino acid arginine starvation efficiently sensitizes cancer cells to canavanine treatment and irradiation, Int. J. Cancer 130 (2012) 2164.
[120] R. Prazeres, F. Glotin, C. Insa, D.A. Jaroszynski, J.M. Ortega, Two- colour operation of a free-electron laser and applications in the mid-infrared, Eur. J. Mass Spectrom. 3 (1998) 87.
[121] L.M. Aleese, A. Simon, T.B. McMahon, J.M. Ortega, D. Scuderi, J. Lemaire, P. Maitre, Mid-IR spectroscopy of protonated leucine methyl ester performed with an FTICR or a Paul type ion-trap, Int. J. Mass Spectrom. 249 (2006) 14.
[122] J.M. Bakker, T. Besson, J. Lemaire, D. Scuderi, P. Maitre, Gas-phase structure of a π-allyl-palladium complex: efficient infrared spectroscopy in a 7 T Fourier transform mass spectrometer, J. Phys. Chem. A 111 (2007) 13415.
[123] J.M. Bakker, R.K. Sinha, T. Besson, M. Brugnara, P. Tosi, J.Y. Salpin, P. Maitre, Tautomerism of uracil probed via infrared spectroscopy of singly hydrated protonated uracil, J. Phys. Chem. A 112 (2008) 12393.
[124] J.K. Martens, G. Berden, C.R. Gebhardt, J. Oomens, Infrared ion spectroscopy in a nodified quadrupole ion trap mass spectrometer at the FELIX free electron laser laboratory, Rev. Sci. Instrum. 87 (2016), http://dx.doi.org/10.1063/1.4964703.
[125] M.J. Deery, S.G. Summerfield, A. Buzy, K.R. Jennings, A mechanism for the loss of 60 u from peptides containing an arginine residue at the C-terminus, J. Am. Soc. Mass Spectrom. 8 (1997) 253.
[126] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision E.01 E. 01, Gaussian, Inc., Wallingford, CT, 2013.
[127] PCModel Serena Software, 2006.
[128] A.D. Becke, Density functional thermochemistry: III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648.
[129] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785.
[130] P.J. Hay, W.R. Wadt, Ab initio effective core potentials for molecular calculations—potentials for the transition-metal atoms Sc to Hg, J. Chem. Phys. 82 (1985) 270.

[131] P.J. Hay, W.R. Wadt, Ab initio effective core potentials for molecular calculations—potentials for K to Au including the outermost core orbitals, J. Chem. Phys. 82 (1985) 299.
[132] W.R. Wadt, P.J. Hay, Ab initio effective core potentials for molecular calculations—potentials for main group elements Na to Bi, J. Chem. Phys. 82 (1985) 284.
[133] T.H. Dunning Jr., P.J. Hay, Modern Theoretical Chemistry, vol. 3, Plenum, New York, 1976.
[134] A.P. Scott, L. Radom, Harmonic vibrational frequencies: an evaluation of Hartree-Fock, Moeller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors, J. Phys. Chem. 100 (1996) 16502.
[135] J.K. Martens, J. Grzetic, G. Berden, J. Oomens, Gas- phase conformations of small polyprolines and their fragment ions by IRMPD spectroscopy, Int. J. Mass Spectrom. (2014), http://dx.doi.org/10.1016/j.ijms.2014.07.027.
[136] A. Baeten, F. De Proft, P. Geerlings, Proton affinity of amino acids: their interpretation with density functional theory-based descriptors, Int. J. Quantum Chem. 60 (1996) 931.
[137] J. Norberg, N. Foloppe, L. Nilsson, Intrinsic relative stabilities of the neutral tautomers of arginine side-chain models, J. Chem. Theory Comput. 1 (2005) 986.
[138] C. Muetterties, A. Drissi Touzani, I. Hardee, K.T. Huynh, J.C. Poutsma, Gas-phase acid-base properties of 1-aminocycloalkane-1-carboxylic acids from the extended kinetic method, Int. J Mass Spectrom. (2014), http://dx.doi.org/10.1016/j.ijms.2014.07.010.
[139] E.P. Hunter, S.G. Lias, Evaluated gas phase basicities and proton affinities of molecules: an update, J. Phys. Chem. Ref. Data 27 (1998) 3.
[140] M. Rozman, Proton affinity of several basic non-standard amino acids, Chem. Phys. Lett. 543 (2012) 50.

THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

pubs.acs.org/JPCC

C–H···O Hydrogen Bonding in Pentamers of Isatin

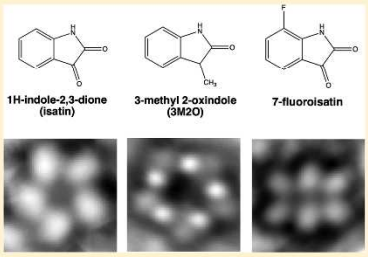
Angela M. Silski,[†] Ryan D. Brown,[†] Jacob P. Petersen,[†] Joseph M. Coman,[†] David A. Turner,[†] Zach M. Smith,[‡] Steven A. Corcelli,[†] John C. Poutsma,[‡] and S. Alex Kandel*[,†]

[†]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States

[‡]Department of Chemistry, The College of William and Mary, Williamsburg, Virginia 23187, United States

Supporting Information

ABSTRACT: Self-assembled monolayers of 1*H*-indole-2,3-dione (isatin), 3-methyl 2-oxindole, and 7-fluoroisatin are observed on the Au(111) surface via scanning tunneling microscopy (STM). We observed that isatin forms pentamers with density functional theory providing support for a cyclic structure stabilized by both N–H···O and C–H···O hydrogen bonds between neighboring molecules. The C–H···O bond is made between the 7-position C–H acting as the hydrogen bond donor and the 3-position carbonyl as the hydrogen bond acceptor, and calculations show that the isatin pentamer structure is 12 kJ/mol more stable than the dimer. When the 3-position carbonyl is removed and replaced with a methyl group (3-methyl 2-oxindole), we observe a monolayer with a mixture of catemer chains and pentameric clusters that are qualitatively different from those of isatin. Pentamer formation is completely broken when the 7-position hydrogen is removed and replaced with fluorine; the monolayer of 7-fluoroisatin is composed of a mixture of close packed ordered domains and hexamer clusters. The role of C–H···O bonding in forming isatin pentamers is supported by electrospray ionization mass spectrometry measurements, which show a propensity for isatin cluster formation, including magic-number isatin pentamers, while 3-methyl 2-oxindole and 7-fluoroisatin show relatively little clustering under the same conditions.



INTRODUCTION

Nanoscaled architectures built from molecular building blocks have a variety of applications such as use in organic semiconductors,[1,2] energy storage and conversion,[3] capacitors,[4] molecular electronics,[5] and host–guest chemistry.[6,7]

Controlled assembly of surface-supported supramolecular structures has been realized by carefully tuning both system energetics[8,9] and noncovalent interactions.[10,11] The hydrogen bond is the strongest and most directional of the noncovalent interactions[12] and is thus a powerful tool in driving molecular self-assembly. The strength of the hydrogen bond can vary greatly with the strongest of O–H···O hydrogen bonds having a binding energy of 100 kJ/mol[13] to a few kJ/mol for weak C–H···O hydrogen bonds.[14] H···A (where A is the hydrogen bond acceptor) distances are typically 1.2–1.5 Å for strong, 1.5–2.2 Å for moderate, and greater than 2.2 Å for weak hydrogen bonds.[12] Strong hydrogen bonds, typically homonuclear, (O–H···O) can be considered mostly covalent in character, while moderate to weak hydrogen bonds are largely electrostatic interactions.[15]

Low-temperature scanning tunneling microscopy (LT-STM), because of its molecular/submolecular resolution, is a useful technique used to image such assemblies of molecules on surfaces and elucidate intra- and intermolecular bonding motifs[16] and, when used in combination with density functional theory (DFT), has been shown to elucidate hydrogen bond directionality.[17,18] Due to the strength of the hydrogen bond, molecular building blocks with strong (O–H···O) to moderate (N–H···O) hydrogen bond interactions are often used as the basis for supramolecular assembly. These hydrogen bonding motifs have resulted in both complex surface-supported structures observed via STM[19−23] and diverse crystalline packing motifs observed via X-ray analysis.[24] In particular, molecules with more than one functional group acting as hydrogen-bond donors and/or acceptors can demonstrate structural polymorphism with many possible geometries both for small molecular clusters and extended crystals.[25−28]

Weaker noncovalent interactions should not be overlooked as insignificant in contributing to such structural polymorphism. Cooperativity of strong hydrogen bonds with weaker interactions is known to play a critical role in increasing hydrogen bond strength,[29] which may lead to unique crystalline packing patterns.[30,31] Many novel self-assembled systems have been realized through the synergy between hydrogen bonding and weaker interactions:[32] surface-supported Sierpiński triangles,[33] guanine quartet networks,[34] and hierarchical supramolecular polymers.[35,36] In particular, the C–H···O hydrogen bond, whose existence initially was met with much debate, is

Received: July 27, 2017
Revised: September 6, 2017
Published: September 15, 2017

DOI: 10.1021/acs.jpcc.7b07444
J. Phys. Chem. C 2017, 121, 21520−21526

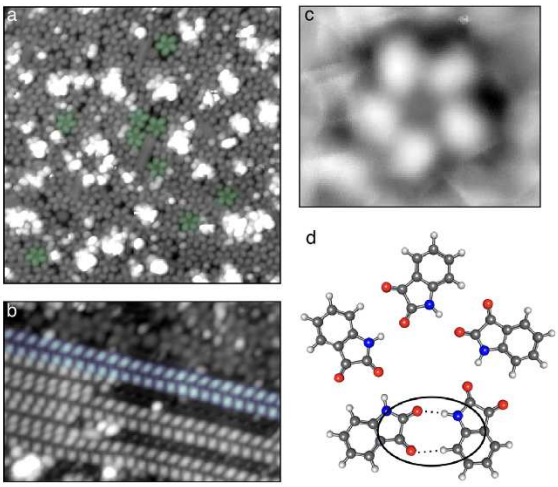


Figure 1. (a) STM image (250 × 250 Å) of isatin vapor deposited on Au(111) surface reveals pentameric structures (representative structures highlighted in green). (b) A 125 × 195 Å image post-thermal annealing where pentamers reorganize to dimer structures. (c) 30 × 25 Å composite image of isatin pentamers from image shown in panel a. (d) Optimized geometry of isatin pentamer with N−H···O and C−H···O hydrogen bond contacts circled.

now well-established as important in small molecule crystal packing,[30,37,38] protein structure,[39,10] crystal engineering,[29,41] and two-dimensional self-assembly.[42,43] The C−H···O hydrogen bond has been reported to have binding energies as high as 14.6 kJ/mol for a parallel β-sheet[44] and 17.6 kJ/mol for an antiparallel β-sheet,[44] while the shortest H···O distance of a C−H···O hydrogen bond has been reported to be 2.07 Å[45]

In this manuscript, we report on the self-assembly of 1*H*-indole-2,3-dione (isatin) on the Au(111) surface. Isatin forms cyclic pentamers stabilized by a combination of N−H···O and C−H···O hydrogen bonds with a per-molecule binding energy that we calculated by DFT to be 12 kJ/mol more stable than the isolated dimer structure. Isatin pentamers are also observed as "magic number" clusters in the gas phase through electrospray ionization mass spectrometry (ESI-MS). In comparison, the molecules 3-methyl 2-oxindole and 7-fluoroisatin have functional group substitutions that preclude C−H···O hydrogen bonding. These two molecules have no magic number clusters in their mass spectra and show significantly different structures on the Au(111) surface.

METHODS

Electrospray Ionization Mass Spectrometry. Mass spectra were obtained in a Thermo Quantum Ultra triple quadrupole instrument equipped with a HESI II heated electrospray source. Twenty millimolar solutions of isatin, 3M2O, and 7-fluoroisatin were directly injected into the instrument at flow rates of 10 μL/min. Q1 resolution was set to 1 Da. Electrospray source conditions were adjusted to minimize cluster dissociation. Typical conditions were: heated electrospray source (30 °C), heated capillary (75 °C), N_2 sheath gas (10 units), ion sweep gas (1 unit), and auxiliary N_2 gas off. Spectra were collected for 10 min with an individual scan time of 3 s. Average mass spectra during the scan are generated in XCalibur 3.0.

Scanning Tunneling Microscopy. Au(111) on mica substrates was cleaned by three rounds of Ar^+ sputtering (20 mA, 0.55 kV, 15 min) and annealing at 350−400 °C (15 min) in a high vacuum sample preparation chamber. The sample was allowed to cool before it was transferred to a high vacuum load-lock chamber. Monolayers of isatin, 3-methyl 2-oxindole, and 7-fluoroisatin were formed through vapor deposition in the load-lock chamber by heating a quartz crucible containing the relevant molecule to 45−50 °C.

The sample was then transferred to an Omicron LT-STM (base pressure 10^{-10} Torr) and cooled to 77 K. In cases where additional annealing was performed, samples were transferred to the sample preparation chamber, heated to 40 °C for 20 min, and allowed to cool to room temperature before being transferred to the STM and cooled to 77 K. All images were acquired with a Pt/Ir tip in constant current mode with a tunneling current of 10 pA and a bias voltage of +1.0 V.

Density Functional Theory. All calculations were done using the Q-Chem software package.[46] Structures were optimized with the Becke three parameter exchange correlation functional with the Lee, Yang, and Parr correlation energy correction (B3LYP). This functional was chosen for its robust parametrization and success with geometric optimizations. The 6-311++G(d,p) basis set was employed with a Lebedev quadrature containing 100 radial shells with 302 angular points. For the ring structures, monomer starting positions were selected such that C_{nh} (where n is the number of monomers in the structure) geometry was approximated. Calculations were performed as isolated gas-phase molecules, and the Au(111) surface was not taken into account. To address basis set superposition error, the Boys and Bernardi counterpoise correction was applied to all systems of two or more molecules.

DOI: 10.1021/acs.jpcc.7b07444
J. Phys. Chem. C 2017, 121, 21520−21526

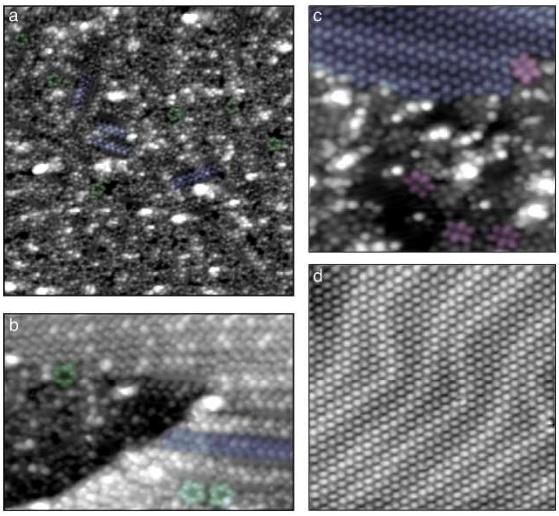


Figure 2. STM images of vapor-deposited monolayers of isatin-related molecules that prevent C−H···O hydrogen bonding. 3M2O images: (a) 350 × 350 Å topography image shows a mixture of catemer structures (purple) and pentamers (green). (b) A 175 × 200 Å close up view of the catemer and pentamer structures. 7-fluoroisatin images: (c) A 158 × 158 Å image, where both hexamers (pink) and ordered close packed molecules are present (purple). (d) A 200 × 200 Å image of close packed molecules, where the periodic change in contrast corresponds to the 22 × $\sqrt{3}$ Au(111) herringbone surface reconstruction.

■ RESULTS AND DISCUSSION

STM Images. Symmetric five-lobed structures are the dominant features observed in STM images of isatin monolayers on Au(111) (Figure 1). We identify these features as pentameric molecular clusters that are chiral upon surface adsorption, similar to those observed for ferrocenecarboxylic acid and indole-2- and indole-3-carboxylic acids.[43,47] The monolayer is composed almost entirely of pentamers with irregular bright features ascribed to the beginnings of multilayer formation.

A significant difference between isatin monolayers and earlier observations of pentamers of indole-2-carboxylic acid and ferrocenecarboxylic acid is that the latter two were prepared through pulse deposition of molecules from solution.[43,47] Pulse deposition creates nonequilibrium conditions through rapid solvent evaporation from aerosolized droplets, and there is evidence that anomalous clusters form in solution under these conditions.[48−50] We also performed solution deposition experiments of isatin on Au(111), which produced a submonolayer-coverage sample that consisted of the same pentameric clusters that are shown in Figure 1 (see Figure S1). However, this was the first time that we observed irregular pentamer formation as a result of vapor deposition of the same molecule. For indole-2-carboxylic acid monolayers, vapor deposition results in an ordered monolayer with a structure resembling that of the compound in the solid state, and there is no evidence of pentamers.[51] There has been a case of pentamer formation via cyclic hydrogen bonding of *trans*-retinoic acid molecules on the Au(111) surface, which was prepared by vapor deposition onto the substrate; however, these pentamers were shown to be surface coverage dependent, where at higher coverages, ordered islands formed.[52] The isatin pentamers do not appear to be either preparation-method or coverage-dependent, as this pentameric structure was seen for both solution- and vapor-deposited samples, which were both low and high coverage substrates, respectively.

The presence of pentamers on the surface upon sublimation is perhaps surprising because there are no bulk crystal structures for isatin that would suggest this geometry, as the solid-state structures of isatin are all based on N−H···O hydrogen-bonded dimers.[53−55] Thermal annealing experiments show that the isatin pentameric structure is metastable, kinetically trapped by the surface growth process during vapor deposition. Supramolecular networks have been shown to be dependent on the energetics (i.e., temperature) of the system[8] and also the surface growth parameters, such as rate of surface diffusion and flux of molecules deposited on the surface.[56] Similarly, *trans*-retinoic acid pentamers were the kinetically grown species at lower temperatures and converted to dimers after annealing.[57]

Figure 1b shows images of the surface after thermal annealing. After annealing, some isatin pentamers remain on the surface, but much of the monolayer is organized into row-like structures that may be either side-by-side dimers or extended (catemer) chains. The disappearance of the pentamer structure and corresponding appearance of the row-like structures after annealing at 40 °C demonstrates that pentamers are metastable and that the thermal energy provided was enough to overcome energy barriers to reorganize into a lower free-energy structure. The known solid state structures of isatin are all based upon N−H···O hydrogen-bonded dimers,[53−55] so

DOI: 10.1021/acs.jpcc.7b07444
J. Phys. Chem. C 2017, 121, 21520−21526

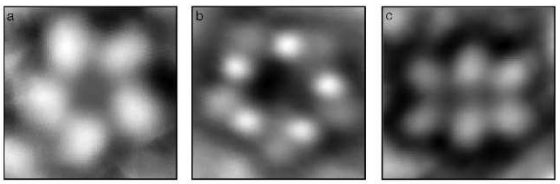


Figure 3. Composite images (20 × 20 Å) of pentamers of (a) isatin and (b) 3M2O and (c) hexamers of 7-fluoroisatin.

the transformation of pentamers to dimers with annealing was expected.

A proposed structure for the isatin pentamer is shown in Figure 1c with the combination of N−H···O and C−H···O hydrogen bonds highlighted. To test the importance of the C−H···O bond, we disrupted it in two ways: we replaced the 3-position carbonyl with a methyl group to get 3-methyl 2-oxindole (3M2O) and replaced the 7-position hydrogen with a fluorine to get 7-fluoroisatin. We found that changing either the hydrogen bond acceptor or donor in these ways significantly changes the monolayer structure.

STM images of the 3M2O monolayer shown in Figures 2a and b demonstrate that the 3M2O monolayer is composed of both catemer chains and pentameric clusters. The catemer chain observed on the 3M2O surface is consistent with the molecule's solid-state crystal structure.[58,59] A monolayer of 7-fluoroisatin is shown in Figure 2c and d. Here, the surface contains domains of ordered close packed molecules in addition to domains of irregular hexameric clusters of molecules. Without the presence of the proton in the 7-position serving as an additional hydrogen bond donor, 7-fluoroisatin does not form pentamers on the surface as was observed for isatin. The substitution of fluorine in the 7-position reinforces that the C−H···O hydrogen bond in our proposed structure (Figure 1c) is critical in stabilizing isatin pentamers.

Composite images of pentamers of isatin, 3M2O, and 7-fluoroisatin are shown in Figure 3, which demonstrate the difference in the isatin and 3M2O pentamer structures and the 7-fluoroisatin hexamer. The contrast within each 3M2O pentamer cluster is significantly different from the isatin pentamer observed. Some of this can be ascribed to changes in the apparent topography of the molecular feature with 3M2O having an additional bright feature that presumably comes from the out-of-plane (and thus topographically higher) methyl group. The 3M2O pentamer structure is much more open than that of isatin, with the axis of each molecule pointing toward the back part of the next molecule. In the isatin pentamer, the axis of each molecule points more to the center of the cluster. While we cannot make a statement on the stability of the 3M2O pentamer in comparison with the isatin pentamer on the basis of our STM data alone, the change in the pentamer structure reveals that removal of the 3-position hydrogen bond acceptor qualitatively changes the bonding motif of the pentamer structure.

Gas-Phase Clusters in ESI. In mass spectrometry "magic number" clusters, or clusters with enhanced stability, can be clearly revealed as peaks with a great relative intensity in comparison with neighboring peaks. Electrospray ionization, because it is a soft ionization technique, allows for the study of clustering behavior of gas-phase ions[60] and has been used to characterize stability of anomalous magic number clusters.[50,61,62] A positive-ion-mode ESI mass spectrum of a 20 mM solution of isatin in acetonitrile is shown in Figure 4a. The mass spectrum reveals peaks $[M_n + H]^+$ peaks $n = 1-8$ as well as small peaks for sodium-ion and ammonium-ion adducts. In electrospray ionization, there is generally a trend of decaying signal intensity with increasing cluster size. Figure 4b is a plot of the natural log of the integrated intensity of each peak as a function of cluster size with all adducts included for each value of n. There is a clear deviation from a simple exponential decay beginning at $n = 5$; notably, the intensity of the pentamer peaks is 4.8 times greater than would be expected based on the intensities of $n = 2-4$. Instrument conditions were tuned to maximize ion transmission and ionic clustering, but qualitatively speaking, this greater than expected relative intensity of the isatin pentamer was consistent at various electrospray conditions (i.e. capillary and cone voltage). While we recognize that the this isatin pentamer peak is not quite as "magic" as the 1,1-ferrocenedicarboxylic acid hexamer[50] or the serine octamer,[61] it does break the trend of exponential decay of signal intensity with increasing cluster size.

Additionally, $n = 6$ and 7 clusters are anomalously intense peaks in ESI-MS; however, these clusters do not appear as surface-supported structures imaged via STM. One possibility is that these are base $n = 5$ clusters with additional isatin molecules attached as adducts. Alternately, it could be possible that these larger clusters are formed in solution conditions but are not stable in 2D adsorption geometries.

ESI data for 3M2O and 7-fluoroisatin are shown in Figures 4c−e. For 3M2O, there is relatively little clustering, and there is a clear exponential falloff from dimers to larger clusters, as represented in Figure 4. For 7-fluoroisatin, only the monomer and dimer peaks are observed. While it is not apparent that the ionic gas-phase clusters produced via ESI are directly comparable to the neutral surface adsorbed molecular clusters that we image via STM, the greater relative intensity of isatin pentamer clusters in its mass spectrum corroborates our observation of pentameric clusters present on a Au(111) surface, regardless of sample preparation method.

Density Functional Theory. Optimized minimum-energy dimer structures were calculated using DFT, where face-to-face positioning of the functional groups allows each N−H to act as a hydrogen-bond donor and each C=O as a hydrogen-bond acceptor, are shown in Figure 5. The per-molecule binding energies are relatively constant across this family of molecules. This is not surprising, as the 3- and 7- position sites are at some distance from the N−H···O bonding region, and distance will reduce the effect of any electrostatic repulsion between O and F atoms. The structures shown in Figure 5 are qualitatively similar to the crystal structures. We expect to see differences with the calculated geometries, which are isolated gas-phase

DOI: 10.1021/acs.jpcc.7b07444
J. Phys. Chem. C 2017, 121, 21520−21526

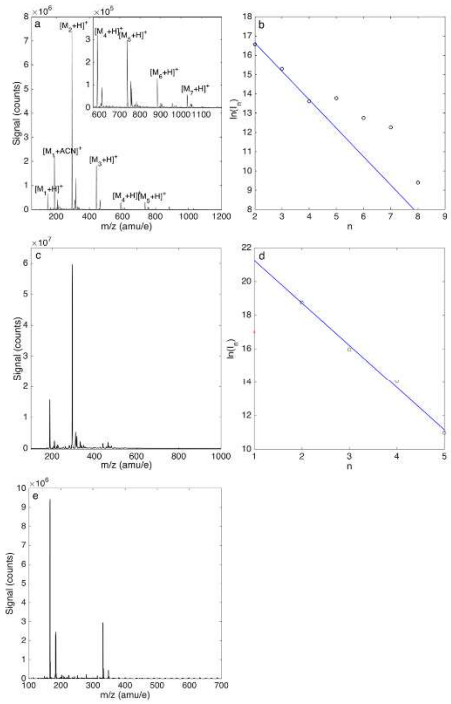


Figure 4. (a) Mass spectrum of 20 mM solution of isatin in acetonitrile, showing $[M_1 + H]^+$ and cluster peaks. (b) Plot of the logarithm of peak intensity, including all adducts, versus cluster size, *n*. The line is fit to $n = 2$–4, and the deviation from the expected exponential decay is apparent for $n = 5$ and larger clusters. ESI mass spectrum of 20 mM solution of 3M2O in acetonitrile. (b) Plot of the natural log of integrated intensities as a function of cluster size illustrates that the signal decays exponentially with increasing cluster size. (c) ESI mass spectrum of 20 mM solution of 3M2O in acetonitrile. (d) Plot of the natural log of integrated intensities as a function of cluster size illustrates that the signal decays exponentially with increasing cluster size. (e) ESI mass spectrum of 20 mM solution of 7-fluoroisatin in acetonitrile; only $[M_1 + H]^+$ and $[M_2 + H]^+$ peaks are present.

dimers, and the crystal structure elucidated by X-ray diffraction. Most notable of these differences is the N–H···O bond distances, which are slightly shorter: 1.88, 1.86, and 1.88 Å calculated versus 1.94, 1.96, and 2.03 Å actual for isatin, 3M2O, and 7-fluoroisatin, respectively.

Optimized minimum-energy cyclic pentamer structures and their corresponding binding energies for the three molecules are also shown in Figure 5. Comparing these three molecules, only the isatin pentamer has a short N–H···O hydrogen bond: 1.87 Å, virtually the same as that calculated for the dimer. The 3M2O N–H···O bond is longer (1.97 Å), and the 7-fluoroisatin bond very long (2.49 Å). For 7-fluoroisatin, C–H···O hydrogen bonding is completely blocked, giving this compound the lowest per-molecule binding energy of them and making the

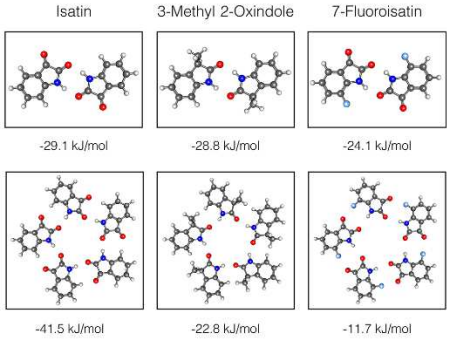


Figure 5. Optimized geometries for dimers and C_5-symmetric pentamers of isatin, 3-methyl 2-oxindole, and 7-fluoroisatin.

pentamer 12.47 kJ/mol less stable than the dimer. For 3M2O, the pentamer and the dimer are of similar stability, and it is interesting to note that the carbonyl groups in the 3M2O pentamer are tilted slightly, possibly to allow some degree of interaction with the 7-position C–H groups. Finally, for isatin, the pentamer structure is calculated to have a per-molecule binding energy of 41.54 kJ/mol, 12.42 kJ/mol more stable than the dimer due to the addition of C–H···O hydrogen bonds. For comparison, the binding energy for N–H···O and C–H···O hydrogen bonds for dipeptides have been calculated to be 30.1 and 11.7 kJ/mol, respectively.[63] Interestingly enough, the sum of these N–H···O and C–H···O binding energies is roughly equivalent to our calculated per-molecule binding energy of the isatin pentamer. Unlike some computational approaches,[64,65] DFT does not produce a breakdown of the percent contribution of the N–H···O and C–H···O hydrogen bond contribution; however, we believe that the presence of the C–H···O hydrogen in the isatin pentamer structure is a critical contributing factor to the exceptional stability of the isatin pentamer. This exceptional stability of the isatin pentamer is consistent with the observation that only isatin has higher-*n* oligomers in its mass spectrum, and 3M2O and 7-fluoroisatin do not.

While we observe pentamers of isatin at monolayer coverage on Au(111), this bonding motif has not been observed in the solid state. The most likely explanation for this is that the stability we calculate for isolated pentamers does not include C–H···O and C–H··· π interactions between neighboring clusters and that these interactions in the extended 3D structure make the dimer the lower-energy form once again. An interesting proposal would be that if isatin could be functionalized, perhaps with bulky substituents at the 4, 5, or 6 positions, these interactions would be disrupted, and such a compound might have a pentamer-based solid-state structure.

■ CONCLUSIONS

Isatin forms metastable symmetric pentamers when self-assembled on Au(111) and similarly shows a preference for the formation of pentameric clusters in electrospray mass spectrometry. Density functional calculations show that by adopting a cyclic structure, isatin molecules in the pentamer can form both N–H···O and C–H···O hydrogen bonds with their neighbors. We focused on the role of C–H···O hydrogen

DOI: 10.1021/acs.jpcc.7b07444
J. Phys. Chem. C 2017, 121, 21520−21526

bonding, where the 7-position C–H is the hydrogen-bond donor and the 3-position carbonyl oxygen is the acceptor. The pentameric structure changes qualitatively when the 3-position carbonyl group is replaced by a methyl to make 3-methyl 2-oxindole and is disrupted completely when the 7-position hydrogen is replaced by a fluorine to make 7-fluoroisatin. Similarly, neither 3-methyl 2-oxindole nor 7-fluoroisatin show any propensity for forming larger molecular clusters in mass spectrometric measurements.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.7b07444.

Images of isatin monolayers prepared by pulsed deposition of isatin solution on Au(111); calculated energies for isatin, 3M2O, and 7-fluoroisatin clusters as a function of cluster size; and Cartesian coordinates for the calculated isatin pentamer (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: skandel@nd.edu.

ORCID

Steven A. Corcelli: 0000-0001-6451-4447

John C. Poutsma: 0000-0002-0085-4079

S. Alex Kandel: 0000-0001-8191-1073

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support for this work was provided by the National Science Foundation (Grant NSF CHE-1507213). We thank the University of Notre Dame Mass Spectrometry Facility for their instrumentation and technical expertise.

■ REFERENCES

(1) Glowacki, E. D.; Irimia-Vladu, M.; Bauer, S.; Sariciftci, N. S. Hydrogen-bonds in molecular solids–From biological systems to organic electronics. *J. Mater. Chem. B* **2013**, *1*, 3742–3753.

(2) Rancatore, B. J.; Mauldin, C. E.; Tung, S.-H.; Wang, C.; Hexemer, A.; Strzalka, J.; Fréchet, J. M. J.; Xu, T. Nanostructured organic semiconductors via directed supramolecular assembly. *ACS Nano* **2010**, *4*, 2721–2729.

(3) Arico, A. S.; Bruce, P.; Scrosati, B.; Tarascon, J.-M.; van Schalkwijk, W. Nanostructured materials for advanced energy conversion and storage devices. *Nat. Mater.* **2005**, *4*, 366–377.

(4) Simon, P.; Gogotsi, Y. Materials for electrochemical capacitors. *Nat. Mater.* **2008**, *7*, 845–854.

(5) Xiang, D.; Wang, X.; Jia, C.; Lee, T.; Guo, X. Molecular-scale electronics: from concept to function. *Chem. Rev.* **2016**, *116*, 4318–4440.

(6) Teyssandier, J.; De Feyter, S.; Mali, K. S. Host-guest chemistry in two-dimensional supramolecular networks. *Chem. Commun.* **2016**, *52*, 11465–11487.

(7) Iritani, K.; Tahara, K.; De Feyter, S.; Tobe, Y. Host-guest chemistry in integrated porous space formed by molecular self-assembly at liquid-solid interfaces. *Langmuir* **2017**, *33*, 4601–4618.

(8) Stannard, A.; Russell, J. C.; Blunt, M. O.; Salesiotis, C.; Giménez-López, M. d. C.; Taleb, N.; Schröder, M.; Champness, N. R.; Garrahan, J. P.; Beton, P. H. Broken symmetry and the variation of critical properties in the phase behaviour of supramolecular rhombus tilings. *Nat. Chem.* **2011**, *4*, 112–117.

(9) Blunt, M. O.; Russell, J. C.; Giménez-López, M. d. C.; Garrahan, J. P.; Lin, X.; Schröder, M.; Champness, N. R.; Beton, P. H. Random tiling and topological defects in a two-dimensional molecular network. *Science* **2008**, *322*, 1077–1081.

(10) Yokoyama, T.; Yokoyama, S.; Kamikado, T. Selective assembly on a surface of supramolecular aggregates with controlled size and shape. *Nature* **2001**, *413*, 619–621.

(11) Hirsch, B. E.; McDonald, K. P.; Flood, A. H.; Tait, S. L. Living on the edge: Tuning supramolecular interactions to design two-dimensional organic crystals near the boundary of two stable structural phases. *J. Chem. Phys.* **2015**, *142*, 101914.

(12) Steiner, T. The hydrogen bond in the solid state. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.

(13) Emsley, J.; Ma, L.; Nyburg, S.; Parkins, A. β-diketone interactions: The structure and properties of 3-(3′,4′,5′-trimethylphenyl)-pentane-2,4-dione, $C_{14}H_{18}O_2$; the hydrogen-bond energy of the enol tautomers of β diketones. *J. Mol. Struct.* **1990**, *240*, 59–67.

(14) Desiraju, G. R.; Steiner, T. The weak hydrogen bond in structural chemistry and biology. *J. Am. Chem. Soc.* **2001**, *123*, 191–192.

(15) Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. Evidence for resonance-assisted hydrogen bonding. Covalent nature of the strong homonuclear hydrogen bond. Study of the O-H–O system by crystal structure correlation methods. *J. Am. Chem. Soc.* **1994**, *116*, 909–915.

(16) Gross, L. Recent advances in submolecular resolution with scanning probe microscopy. *Nat. Chem.* **2011**, *3*, 273–278.

(17) Lawton, T. J.; Carrasco, J.; Baber, A. E.; Michaelides, A.; Sykes, E. C. H. Visualization of hydrogen bonding and associated chirality in methanol hexamers. *Phys. Rev. Lett.* **2011**, *107*, 256101.

(18) Guo, J.; Meng, X.; Chen, J.; Peng, J.; Sheng, J.; Li, X.-Z.; Xu, L.; Shi, J.-R.; Wang, E.; Jiang, Y. Real-space imaging of interfacial water with submolecular resolution. *Nat. Mater.* **2014**, *13*, 184–189.

(19) Theobald, J. A.; Oxtoby, N. S.; Phillips, M. A.; Champness, N. R.; Beton, P. H. Controlling molecular deposition and layer structure with supramolecular surface assemblies. *Nature* **2003**, *424*, 1029–1031.

(20) Slater Nee Phillips, A. G.; Beton, P. H.; Champness, N. R. Two-dimensional supramolecular chemistry on surfaces. *Chem. Sci.* **2011**, *2*, 1440–1448.

(21) Slater Nee Phillips, A. G.; Perdiagao, L.; Beton, P. H.; Champness, N. R. Surface-based supramolecular chemistry using hydrogen bonds. *Acc. Chem. Res.* **2014**, *47*, 3417–3427.

(22) De Feyter, S.; De Schryver, F. C. Two-dimensional supramolecular self-assembly probed by scanning tunneling microscopy. *Chem. Soc. Rev.* **2003**, *32*, 139.

(23) Bartels, L. Tailoring molecular layers at metal surfaces. *Nat. Chem.* **2010**, *2*, 87–95.

(24) Ivasenko, O.; Perepichka, D. F. Mastering fundamentals of supramolecular design with carboxylic acids. Common lessons from X-ray crystallography and scanning tunneling microscopy. *Chem. Soc. Rev.* **2011**, *40*, 191–206.

(25) Xantheas, S. S. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.* **2000**, *258*, 225–231.

(26) Saenger, W.; Lindner, K. OH clusters with homodromic circular arrangement of hydrogen bonds. *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 398–399.

(27) Ahn, S.; Guo, F.; Kariuki, B. M.; Harris, K. D. M. Abundant polymorphism in a system with multiple hydrogen-bonding opportunities: Oxalyl dihydrazide. *J. Am. Chem. Soc.* **2006**, *128*, 8441–8452.

(28) Chen, S.; Guzei, I. A.; Yu, L. New polymorphs of ROY and new record for coexisting polymorphs of solved structures. *J. Am. Chem. Soc.* **2005**, *127*, 9881–9885.

(29) Desiraju, G. R. The C–H···O hydrogen bond: Structural implications and supramolecular design. *Acc. Chem. Res.* **1996**, *29*, 441–449.

(30) Sharma, C. K.; Desiraju, G. R. C–H···O hydrogen bond patterns in crystalline nitro compounds: studies in solid-state

molecular recognition. *J. Chem. Soc., Perkin Trans. 2* **1994**, *2*, 2345−2352.
(31) Braga, D.; Grepioni, F.; Desiraju, G. R. Crystal engineering and organometallic architecture. *Chem. Rev.* **1998**, *98*, 1375−1405.
(32) Langner, A.; Tait, S.; Lin, N.; Chandrasekar, R.; Ruben, M.; Kern, K. Ordering and stabilization of metal−organic coordination chains by hierarchical assembly through hydrogen bonding at a surface. *Angew. Chem., Int. Ed.* **2008**, *47*, 8835−8838.
(33) Shang, J.; Wang, Y.; Chen, M.; Dai, J.; Zhou, X.; Kuttner, J.; Hilt, G.; Shao, X.; Gottfried, J. M.; Wu, K. Assembling molecular Sierpiński triangle fractals. *Nat. Chem.* **2015**, *7*, 389−393.
(34) Otero, R.; Schöck, M.; Molina, L. M.; Lægsgaard, E.; Stensgaard, L.; Hammer, B.; Besenbacher, F. Guanine quartet networks stabilized by cooperative hydrogen bonds. *Angew. Chem., Int. Ed.* **2005**, *44*, 2270−2275.
(35) Rest, C.; Kandanelli, R.; Fernández, G. Strategies to create hierarchical self-assembled structures via cooperative non-covalent interactions. *Chem. Soc. Rev.* **2015**, *44*, 2543−2572.
(36) Yan, X.; Li, S.; Pollock, J. B.; Cook, T. R.; Chen, J.; Zhang, Y.; Ji, X.; Yu, Y.; Huang, F.; Stang, P. J. Supramolecular polymers with tunable topologies via hierarchical coordination-driven self-assembly and hydrogen bonding interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 15585−90.
(37) Taylor, R.; Kennard, O. Crystallographic evidence for the existence of C−H···O, C−H···N and C−H···Cl hydrogen bonds. *J. Am. Chem. Soc.* **1982**, *104*, 5063−5070.
(38) Steiner, T.; Saenger, W. Geometry of C−H···O hydrogen in carbohydrate crystal structures. Analysis of neutron diffraction data. *J. Am. Chem. Soc.* **1992**, *114*, 10146−10154.
(39) Derewenda, Z. S.; Lee, L.; Derewenda, U. The occurrence of C−H···O hydrogen bonds in proteins. *J. Mol. Biol.* **1995**, *252*, 248−262.
(40) Ash, E. L.; Sudmeier, J. L.; Day, R. M.; Vincent, M.; Torchilin, E. V.; Haddad, K. C.; Bradshaw, E. M.; Sanford, D. G.; Bachovchin, W. W. Unusual ^{1}H NMR chemical shifts support (His) C$^{\varepsilon 1}$−H···O = C H-bond: Proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10371−10376.
(41) Desiraju, G. R. C−H···O and other weak hydrogen bonds. From crystal engineering to virtual screening. *Chem. Commun.* **2005**, *24*, 2995−3001.
(42) Miao, K.; Hu, Y.; Zha, B.; Xu, L.; Dong, M.; Miao, X.; Deng, W. Polymorphic self-sssemblies of 2,7-Bis(decyloxy)-9-fluorenone at the solid/gas interface: role of C−H···O═C hydrogen bond. *J. Phys. Chem. C* **2017**, *121*, 3947−3957.
(43) Wasio, N. A.; Quardokus, R. C.; Forrest, R. P.; Lent, C. S.; Corcelli, S. A.; Christie, J. A.; Henderson, K. W.; Kandel, S. A. Self-assembly of hydrogen-bonded two-dimensional quasicrystals. *Nature* **2014**, *507*, 86−89.
(44) Scheiner, S. Weak H-bonds. Comparisons of C−H···O to N−H···O in proteins and P−H···N to direct P···N interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13860−13872.
(45) Allen, F. H.; Howard, J. A. K.; Hoy, V. J.; Desiraju, G. R.; Reddy, D. S.; Wilson, C. C. First neutron diffraction analysis of an O-H··· π hydrogen bond: 2-ethynyladamantan-2-ol. *J. Am. Chem. Soc.* **1996**, *118*, 4081−4084.
(46) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172−3191.
(47) Wasio, N. A.; Quardokus, R. C.; Brown, R. D.; Forrest, R. P.; Lent, C. S.; Corcelli, S. A.; Christie, J. A.; Henderson, K. W.; Kandel, S. A. Cyclic hydrogen bonding in indole carboxylic acid clusters. *J. Phys. Chem. C* **2015**, *119*, 21011−21017.
(48) Quardokus, R. C.; Wasio, N. A.; Brown, R. D.; Christie, J. A.; Henderson, K. W.; Forrest, R. P.; Lent, C. S.; Corcelli, S. A.; Kandel, S. A. Hydrogen-bonded clusters of 1,1′-ferrocenedicarboxylic acid on Au(111) are initially formed in solution. *J. Chem. Phys.* **2015**, *142*, 101927.
(49) Shah, V. B.; Biswas, P. Aerosolized droplet mediated self-assembly of photosynthetic pigment analogues and deposition onto substrates. *ACS Nano* **2014**, *8*, 1429−1438.
(50) Brown, R. D.; Coman, J. M.; Christie, J. A.; Forrest, R. P.; Lent, C. S.; Corcelli, S. A.; Henderson, K. W.; Kandel, S. A. Evolution of metastable clusters into ordered structures for 1,1′-ferrocenedicarboxylic acid on the Au(111) surface. *J. Phys. Chem. C* **2017**, *121*, 6191−6198.
(51) De Marchi, F.; Cui, D.; Lipton-Duffin, J.; Santato, C.; MacLeod, J. M.; Rosei, F. Self-assembly of indole-2-carboxylic acid at graphite and gold surfaces. *J. Chem. Phys.* **2015**, *142*, 101923.
(52) Karan, S.; Wang, Y.; Robles, R.; Lorente, N.; Berndt, R. Surface-supported supramolecular pentamers. *J. Am. Chem. Soc.* **2013**, *135*, 14004−14007.
(53) Palenik, G. J.; Koziol, A. E.; Katritzky, A. R.; Fan, W.-Q. Nonbonded interactions. The influence of lone pair repulsions on bond lengths. *J. Chem. Soc., Chem. Commun.* **1990**, *9*, 715−716.
(54) Palmer, M. H.; Blake, A. J.; Gould, R. ^{14}N nuclear quadropole coupling in cyclic amides and thioamides. Ab Initio simulations of the solid state environment as interpretation of the NQR spectra of 2-pyridinone, isatin and benzothiazolezone. A new X-ray structure for isatin. *Chem. Phys.* **1987**, *115*, 219−227.
(55) Goldschmidt, G. H.; Llewellyn, F. J. The crystalline structure of isatin. *Acta Crystallogr.* **1950**, *3*, 294−305.
(56) Barth, J. V.; Costantini, G.; Kern, K. Engineering atomic and molecular nanostructures at surfaces. *Nature* **2005**, *437*, 671−679.
(57) Li, C.; Li, N.; Liu, L.; Zhang, Y.; Yuan, C.; Peng, L.; Hou, S.; Wang, Y. Kinetically controlled hierarchical self-assemblies of all-trans-retinoic acid on Au(111). *Chem. Commun.* **2017**, *53*, 2252−2255.
(58) Lipkowski, J.; Luboradzki, R.; Stefaniak, L.; Wojick, J. X-ray diffraction study of some indolinones. *J. Chem. Crystallogr.* **1995**, *25*, 299−308.
(59) Hachuła, B.; Flakus, H. T.; Polasz, A. Temperature, H/D isotopic and Davydov-splitting effects in the polarized IR spectra of hydrogen bond chain systems: 1,2,4-triazole and 3-methyl-2-oxindole crystals. *Spectrochim. Acta, Part A* **2014**, *120*, 287−296.
(60) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64−70.
(61) Cooks, R. G.; Zhang, D. X.; Koch, K. J.; Gozzo, F. C.; Eberlin, M. N. Chiroselective self-directed octamerization of serine: Implications for homochirogenesis. *Anal. Chem.* **2001**, *73*, 3646−3655.
(62) Gozzo, F. C.; Santos, L. S.; Augusti, R.; Consorti, C. S.; Dupont, J.; Eberlin, M. N. Gaseous supramolecules of imidazolium ionic liquids: "magic" numbers and intrinsic strengths of hydrogen bonds. *Chem. - Eur. J.* **2004**, *10*, 6187−6193.
(63) Scheiner, S. Relative strengths of N-H···O and C−H···O hydrogen bonds between polypeptide chain segments. *J. Phys. Chem. B* **2005**, *109*, 16132−16141.
(64) Espinosa, E.; Molins, E.; Lecomte, C. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.* **1998**, *285*, 170−173.
(65) Osmialowski, B.; Kolehmainen, E.; Valkonen, A.; Kowalska, M.; Ikonen, S. The influence of CH bond polarization on the self-association of 2-acylaminopyrimidines by NH/CH···O/N interactions: XRD, NMR, DFT, and AIM study. *Struct. Chem.* **2013**, *24*, 2203−2209.

DOI: 10.1021/acs.jpcc.7b07444
J. Phys. Chem. C 2017, 121, 21520−21526